Principles of Electricity

applied to

Telephone and Telegraph Work

1961 Edition

Principles of Electricity

applied to

Telephone and Telegraph Work

Spirit of Communication

PREFACE

THIS book has appeared in a number of editions during the past four decades, the latest prior edition being dated 1953. Its purpose has remained the same throughout the period—namely, to serve as a basic test on elementary electrical principles for use in the technical training of employees in the Long Lines Department of the American Telephone and Telegraph Company.

Although truly fundamental principles do not change with time, the rapid growth and development of the communications art over the years has necessitated the incorporation of numerous new or changed illustrative applications in each succeeding edition. This, of course, remains true for the present volume. In addition, the introduction of new technologies and new instrumentalities of first importance has made it desirable to include much new material in such fields as semi-conductor theory and the transmission of high-frequency electromagnetic waves. This, in turn, has required the introduction of certain basic concepts and principles not dealt with in earlier editions, as well as numerous examples to illustrate their applications in practice.

The variety and scope of the subjects covered in the book make it necessarily rather voluminous even though considerable material included in previous editions has been omitted in this. Every effort has nevertheless been made to treat each subject taken up as briefly as is consistent with a reasonably adequate presentation of the related theory and fields of application.

It is important for the reader to recognize that the book is not, and is not intended to be, a comprehensive treatise embracing the entire field of electrical communications. Its subject is electrical theory. Such descriptions of communication equipment and circuits as are included were selected primarily to illustrate practical applications and many of them have been somewhat arbitrarily chosen. The objective has been only to cover the essential general principles of basic electrical theory and to illustrate each principle briefly by one or more of its significant applications in the communications field.

The use of higher mathematics is avoided entirely in this book, and even the more elementary branches are employed as sparingly as possible. A general knowledge on the part of the reader is assumed of only those branches of mathematics ordinarily taught in High Schools, including Algebra, Geometry, Logarithms and Trigonometry. In addition, it has been thought desirable, in the Chapters dealing with the solution of alternating-current networks and with wire transmission theory, to make some use of simple Vector Notation. This may involve the introduction of certain mathematical concepts not familiar to all readers, but it is believed that the great simplification that may be effected with this convenient mathematical tool will more than justify any additional study time that the reader may find needed to master its practical use.

A general knowledge of elementary Physics and Mechanics is also assumed. The first Chapter of the text, however, reviews very briefly some of the fundamental physical principles that are particularly applicable to various subjects discussed in later Chapters.

A word of caution is perhaps needed regarding the use of the circuit drawings, tables, and

other statistical data included at various points in the text. The circuit drawings are presented primarily as a means of illustrating the principles under discussion. Although they may be reasonably representative of actual practice, they may or may not conform in detail with any situation familiar to the reader. Similarly, the tables and other data represent the best information available at this time, but they are subject to change and are not intended as a substitute for data issued in current formal instructions.

Previous editions of this text have also been signed by L. S. Crosby, General Personnel Supervisor, who is now retired. His major contributions to the original planning and content of the work are cordially acknowledged.

C. F. MYERS
Supervisor of Instruction

32 Avenue of the Americas,
New York, N. Y.,
June, 1961

TABLE OF CONTENTS

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

TABLES

Principles of Electricity
applied to
Telephone and Telegraph Work

CHAPTER 1

BASIC PHYSICAL CONCEPTS

1.1 Introduction

Electricity is an agency of remarkable flexibility, with almost unlimited applications. By means of electricity, energy available at one place, in the form, for example, of burning coal or falling water may be transferred to another place perhaps many miles away, and utilized there in any convenient form, such as heat, light, or mechanical work. Electricity may be used to change the physical or chemical state of substances, and it may be employed to control processes, to check quality, and even to perform mathematical computations. Through the electron, basic unit of electricity, we may probe the physical and biological worlds, and study hidden secrets of the universe.

The power engineer is able, by applying his knowledge of the behavior of electricity, to accomplish the transformation and transmission of large amounts of energy that may be needed for many and various purposes; and the generation, flow and delivery of this energy also can be controlled electrically. The communications engineer, by similarly applying his knowledge of the same basic principles, is able to transmit intelligence in a variety of forms over great distances and without loss of meanings. The source may be a human voice, a musical instrument, printed page or picture, and it can be reproduced in like form, completely recognizable and intelligible, wherever needed.

Since this book is prepared primarily for telephone people, we shall discuss these basic principles with reference, naturally, to applications in the communications field. This will mean, in general, that we will be interested mostly in small values of electrical energy, and in the transmission of such energy over relatively great distances. This interest will embrace both direct current and alternating current concepts and in the underlying alternating current phenomena, the scope will cover a wide range of frequencies, including energy flowing over wires and energy radiated through space.

There are various possible approaches to beginning the study of electricity. In this book, we shall adhere to the conventional, classical approach, beginning with the study of direct current and direct-current circuits, and proceeding from there to alternating current theory, with particular emphasis in the later Chapters on electronics and high-frequency phenomena.

Before beginning the study of electric circuits and the electrical transmission of energy, however, we shall consider briefly, in this Chapter, some of the more fundamental concepts of the physics of electricity.

1.2 Electric Charge

The earliest recorded recognition of electrical phenomena dates back several thousand years when it was discovered that rubbing a piece of amber with woolen cloth would cause it to attract other light bodies, such as bits of paper. The amber had thus acquired a property by virtue of which it exerted a certain force on other materials in its vicinity. This property became known as an electric charge. Then it was discovered that a similar phenomenon occurred when glass was rubbed with silk. Subsequent experimentation demonstrated, however, that the charge on the glass was opposite in sign from the charge on the amber. To distinguish them, the former was labeled "positive charge" and the latter "negative charge". In both cases the magnitude of the charge depended upon the extent of the surface rubbed and the intensity of the friction.

It was also found that two like bodies, such as two pieces of glass or amber, when so treated would exert a repelling force on each other. In other words, positive charges repel each other and negative charges repel each other. On the other hand, a force of attraction was found to exist between a positively charged substance and a negatively charged substance. These attracting and repelling forces although small are millions of times larger than the force of attraction exerted by gravity between similar uncharged bodies.

The "practical" electrical unit for measurement of these electric charges is called the cou-

physical reality but the concept is so convenient that the line of electric force or induction is generally used as a basic unit of measurement. The strength of a field at any point may then be stated as the number of lines per square centimeter, where each line is considered as extending between unit charges of opposite sign.
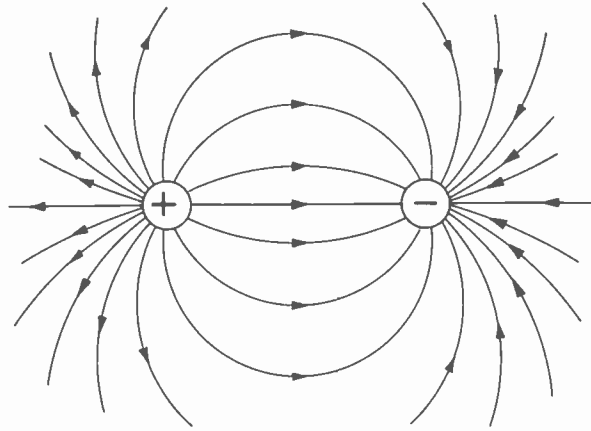
1.5 The Magnetic Field

Also of long standing in the classical theory of electricity is the concept of the magnetic field. This field, as discussed more extensively in Chapter 4, represents the forces to be found in the vicinity of magnets. Like the electric field, its intensity at any point is directly proportional to the strength of the magnetic poles to which the field is due, and inversely proportional to the square of the distance from the poles. This field is similarly pictured by **lines of magnetic force or induction** and its intensity is measured in terms of number of lines of magnetic force per square centimeter. Unlike the electric field, however, magnetic lines of force are not thought of as terminating but as always forming closed loops.

A magnetic field is not only present in the space within and surrounding any magnet but, as is pointed out in Chapter 4, such a field may be detected in the space surrounding any conductor in which electric current is flowing. This phenomenon seems to require the introduction of a new concept, since there is no apparent identity between a static field produced by a magnet and a field associated with current flow in a conductor

and having nothing to do with any magnets or magnetic materials. Yet the two fields appear to have precisely the same characteristics.

Modern electrical theory resolves this problem by considering the magnetic field as a result or accompaniment of a moving electric field. Thus, current flow in a conductor is a movement of electrons, which are tiny electric charges. These charges have accompanying electric fields which of course must move as the electrons move. Extending this idea further, the magnetic field of a magnet is ascribed to the spinning movements of electrons in the magnetic material. From this view, the magnetic field may even be thought of as merely an aspect of the moving electric field. Nevertheless, the concept of the magnetic field as such, and of lines of magnetic force, continues to be very useful in the practical analysis of many electrical phenomena, and it need not be discarded.

1.6 Electro-Magnetic Fields

Under the field concepts discussed above, either an electric field or a magnetic field may exist alone only under static conditions. Under dynamic conditions, where either type of field is changing or moving, the total field must be electro-magnetic— i.e., both an electric field and a magnetic field must exist simultaneously. This is a necessary result of the fact that the moving electric field is always accompanied by a magnetic field, and vice versa. In dealing with electro-dynamic phenomena, accordingly, it is reasonable in many situations to consider only a single composite field which may be designated electro-magnetic. There are often distinct advantages, however, in viewing certain dynamic phenomena on the basis of coexistent electric and magnetic fields, each with its force lines always at right angles to those of the other.

In our study of direct currents and direct-current circuits in the immediately following Chapters, we shall not be much concerned with fields other than static magnetic fields. As we go into the study of high-frequency alternating currents and the behavior of various electronic devices, however, we shall find that dynamic electric and magnetic fields will assume a place of increasing importance—particularly in connection with radio transmission and other transmission at high and super-high frequencies.

CHAPTER 2

DIRECT CURRENT AND DIRECT-CURRENT CIRCUITS

2.1 Direct Current Flow

In discussing electrons in the preceding Chapter it was pointed out that the movement of a stream of electrons through a conductor constitutes an electric current. Where such a flow of electrons is continuous in one direction it is called a **direct current**. It has long been customary to begin the practical study of electricity with an analysis of the behavior of such direct currents, primarily because the phenomena involved are relatively easy to deal with. We shall conform to this well-established study procedure in the beginning of this book although we shall later have to consider in some detail the behavior of currents which reverse their direction of flow at periodic intervals. These latter are called **alternating currents**.

2.2 Electric Pressure or Electromotive Force

In order to establish a flow of current through a conductor it is necessary that there be connected in the circuit a source of electric pressure which sets up what is known as an **electromotive force**. When this condition is met the flow of electricity through a circuit is analogous in many respects to the flow of water through a closed system of pipes. Figure 2-1 shows a simple electric circuit consisting of a battery connected to a resistor *ab*. Figure 2-2 shows a simple water circulating system. In the water mechanism, the pump creates a difference in pressure between the points *a* and *b*. This difference in pressure, or "pressure head", will cause water to flow from the outlet pipe *a*, through the small pipe to the flow meter, and return to
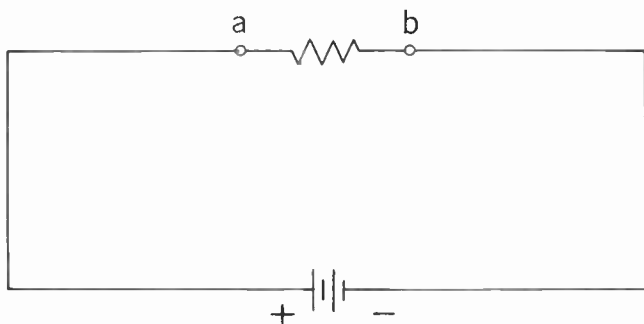
the low pressure side of the pump at *b*. The amount of water that will flow will depend upon this difference in pressure and upon the nature of the small pipe. In the electric circuit, the battery supplies the electric pressure or electromotive force which causes electricity to flow from the "high potential" side of the battery. The amount of electricity that will flow depends upon this electromotive force and the nature of the resistor.
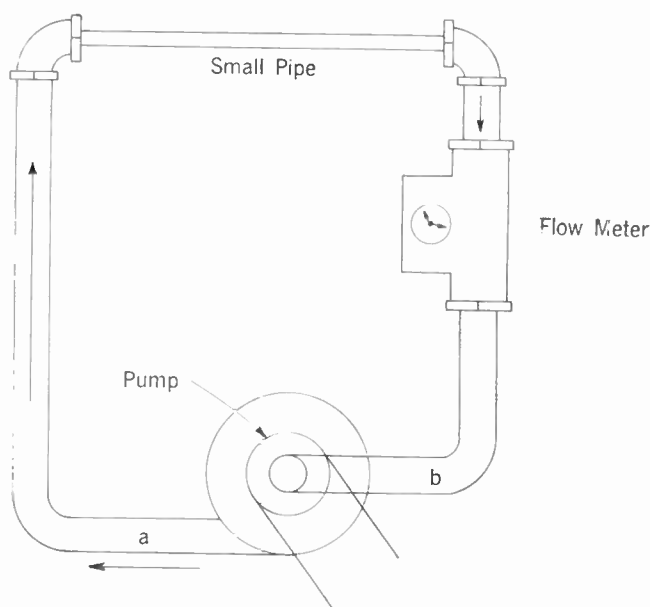
FIG. 2-2 WATER CIRCULATING SYSTEM ANALOGOUS TO SIMPLE ELECTRIC CIRCUIT

If a differential pressure gage were connected between the points *a* and *b* in the water system, it would register the difference in water pressure in some suitable unit such as "difference of head in feet". The electromotive force of the electric circuit, on the other hand, is measured in terms of a unit called the **volt**.

Since a source of electromotive force is by definition capable of exerting electric pressure on any external circuit connected to its terminals, it is customary to say that the "electric potential" of the positive terminal is higher than that of the negative terminal. The difference is the electromotive force of the battery or other source. In a closed circuit, such as shown in Figure 2-3, the

FIG. 2-1 SIMPLE ELECTRIC CIRCUIT

[5]
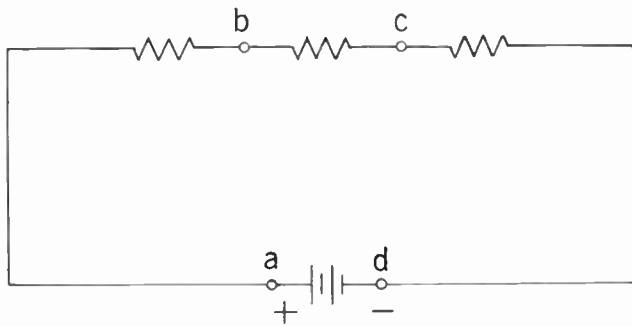
FIGURE 2-3

potential at point *b* will be lower than that at point *a*, that at *c* will be lower than at point *b* and so on around the circuit. Thus we say that there is a **potential drop** from *a* to *b* that may be measured in volts, the magnitude of which depends on the resistance of the conductor between *a* and *b*. Similarly there is another potential drop between *b* and *c*, and *c* and *d*.

2.3 Resistance

In Figure 2-2, if the small pipe is made longer the flow of water will be decreased although the pump maintains a constant difference in pressure between the points *a* and *b*. Also, if the small pipe is decreased in size the flow of water will likewise be decreased. Though there is no simple unit for measuring this resistance to flow of water in a pipe, it is analogous to an electric resistance in many respects. The resistance offered to a flow of water in a pipe is due primarily to friction between the moving molecules of water and the inner walls of the pipe. Resistance to the flow of electric current in a conductor is considered to be due to the multitudinous interactions or collisions between the moving electrons and the atoms of the conducting material. The practical unit of electric resistance is called the **ohm**, which by an act of the U. S. Congress, approved July 21, 1950, is defined as being **equal to one thousand million units (10^9) of resistance of the centimeter-gram-second (cgs) system of electromagnetic units.**

Note: As defined in standard textbooks on physics, the cgs electromagnetic units are derived from the fundamental mechanical units of **length, mass and time** according to the basic principles of electromagnetism, with the value of the permeability of free space arbitrarily chosen as **unity**.

2.4 Current

In our water circulating mechanism we can describe the "rate of flow", or the current, as the amount of water being circulated in gallons per second. The practical unit of electric current is the **ampere** which is defined by law as being equal **to one-tenth of the unit of current of the centimeter-gram-second system of electromagnetic units.**

2.5 The Volt

The **volt** is the practical unit of electric pressure. It is **the electromotive force that, steadily applied to a conductor whose resistance is one ohm, will produce a current of one ampere.**

2.6 Open Circuits

The electric circuits shown thus far indicate no means of interrupting the flow of electricity. For the same reason that any water system should be equipped with valves or other devices for starting and stopping the flow of water, switches, push-buttons, keys, etc. are used for opening and closing electric circuits. Figure 2-4 shows a circuit opened by means of a switch. Its metallic continuity is interrupted by the switch and when so interrupted there is no flow of electricity. This protects the source of electromotive force against unnecessary losses, since when the circuit is open it cannot absorb any energy.

2.7 Electrical Symbols and Circuit Conventions

In the foregoing circuit diagrams we have represented the battery with a long and a short line, a resistance by a wavy line, connecting wires by straight plain lines, and connections between the wires and the battery or the wires and the resistances by small circles. These are circuit conven-

FIGURE 2-4

[6]
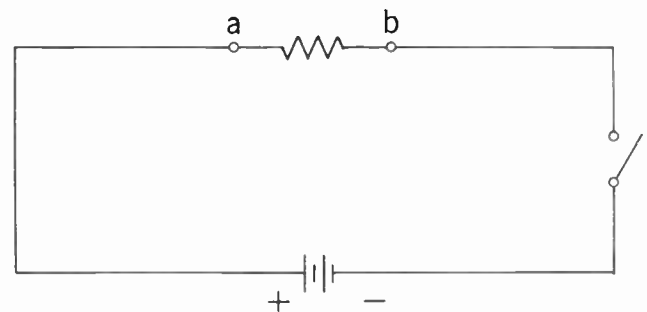
tions. Thus, Figure 2-5 illustrates an actual door bell circuit and Figure 2-6 shows the electrical properties of the same circuit drawn in accordance with standard electrical conventions. There are many such conventions and different ones are used for different purposes. For example, on drawings which are to guide the electrical installer when connecting wires to various units of apparatus, a somewhat different set of conventions is used than on drawings to illustrate a circuit's theory of operation. Figure 2-7 shows a few simple conventions that should be learned at this time.

In addition to the circuit conventions used in illustrating the theory of electrical circuits by diagrams, certain symbols are necessary for representing electrical quantities in simple mathematical equations. Table I gives standard symbols for electrical quantities. It is necessary to learn now those applying to the quantities we have defined. The table can later be referred to for other quantities treated.

2.8 Ohm's Law

A German physicist named George Simon Ohm was the first to discover the relationship between current, electromotive force and resistance. The discovery is called **Ohm's Law** and simply expressed is—**that for any circuit or part of a circuit under consideration the current in amperes is equal to the electromotive force in volts divided by the resistance in ohms.**
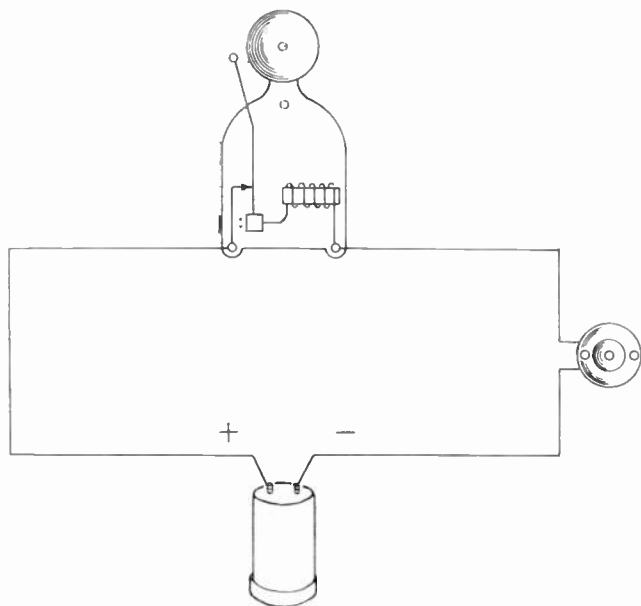
Fig. 2-5 Wiring of Door-bell Circuit
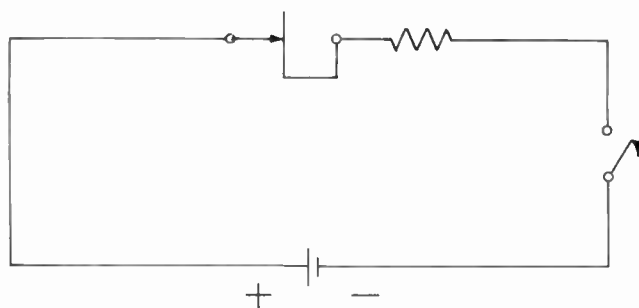
Fig. 2-6 Conventional Drawing for Circuit of Fig. 2-5

This law, mathematically expressed, is as follows:

$$\text{Current} = \frac{\text{Electromotive Force}}{\text{Resistance}}$$

If in the above expression we substitute the proper symbols for current, electromotive force and resistance we have the following equation:

$$I = \frac{E}{R} \qquad (2:1)$$

This is the equation for Ohm's Law. It is perhaps the most important one in all electrical work. It may be expressed in other forms, but when expressed as shown, permits us to calculate the current that may be expected in any circuit when we know the voltage of the source of electromotive force and when we know the resistance connected to this source in ohms.

Example: In Figure 2-8 if the electromotive force of the battery is 24 volts and the resistance of the lamp connected to it is 112 ohms, what will be the value of the current flowing through the lamp when the circuit is closed?

Solution: $E = 24$

$R = 112$

$I = \frac{E}{R} = \frac{24}{112} = .21$ ampere, ans.

2.9 Other Ways of Expressing Ohm's Law

Equation (2:1) states that the current is equal to the electromotive force divided by the resistance; then by simple algebra the electromotive force must be equal to the current multiplied by the resistance, or the equation may be expressed—

$$E = IR \qquad (2:2)$$

From this equation we may find the electromotive

Crossed   Connected   Grounded
        WIRES

Single Cell   24 - Volt
        BATTERIES

RESISTOR

INDUCTOR

Fixed   Variable
    CAPACITOR

VOLTMETER

AMMETER

Switchboard   Resistance
        LAMPS

FUSE

TELEPHONE RECEIVERS

TELEPHONE
TRANSMITTER

RINGER

Locking Lever - Type

Non-locking Lever - Type
        KEYS

Regular Type   Twin Type
        JACKS

2 - Conductor   Twin Type
        PLUGS

Single wound
with make contact
        RELAY

Double wound with break
& make - before - break contact
        RELAY

IN

Retardation   Repeating
        COILS

TRANSFORMER

AMPLIFIER

ANTENNA

METALLIC RECTIFIER

Fig. 2-7 Conventions Commonly Used in Telephone Circuit Drawings

[8]

FIGURE 2-8

force acting in any circuit if we know the resistance and the current.

Example: In Figure 2-6 the resistance of the door bell winding is 4 ohms. If during the instant the circuit is closed the current is .2 ampere, what is the voltage of the dry cell?

Solution: $R = 4$

$I = .2$

$E = IR = .2 \times 4 = .8$ volt, ans.

The third case is one where current and electromotive force are known and it is desired to find the resistance. Ohm's Law may likewise be stated to cover these conditions. If the electromotive force is equal to the resistance multiplied by the current, the resistance must be equal to the electromotive force divided by the current or, algebraically expressed—

$$R = \frac{E}{I} \qquad (2:3)$$

Example: What is the resistance connected between the points a and b in Figure 2-1 if the voltage of the battery is 1.3 volts and the current is .5 ampere?

Solution: $E = 1.3$ volts

$I = .5$ ampere

$R = \frac{E}{I} = \frac{1.3}{.5} = 2.6$ ohms, ans.

2.10 Potential Differences in a Closed Circuit

We have spoken of how the differential pressure gage may measure the difference in pressure head of the two sides of the water pump shown by Figure 2-2. The electrical instrument used for measuring the electric pressure of a source of electromotive force, or the potential dif-

TABLE I

SYMBOLS USED IN ELECTRICAL WORK

I Current in Amperes
E Electromotive Force in Volts
R Resistance in Ohms
P Power in Watts
Q Quantity in Coulombs
V Potential Drop in Volts
G Conductance in Mhos (is reciprocal of R)
T Time in Seconds
C Capacitance in Farads
L Inductance in Henrys
f Frequency in Cycles per Second
X Reactance in Ohms
Z Impedance in Ohms
Y Admittance in Mhos

ference between any two points in a circuit, is called the voltmeter.

Figure 2-9 shows a voltmeter being used to measure the voltage of a dry cell on an open circuit. Figure 2-10 shows the voltmeter connected to measure the voltage of a source of electromotive force in a closed circuit. In this case we have a simple circuit with three resistors in series. If the voltmeter is connected across the points a and b as shown in Figure 2-11, which represents the same circuit as Figure 2-10, its reading will be lower than when connected
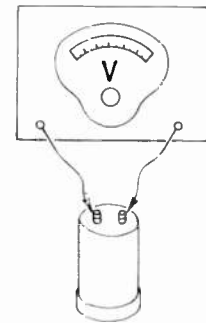
FIGURE 2-9

across the battery. Moreover if the voltmeter is connected across the resistors b and c, and c and d, the three readings, that is, the readings across a and b, b and c, and c and d when added together, will be equal to the voltage of the battery
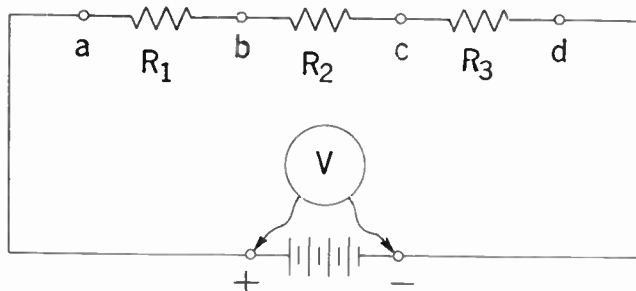
FIG. 2-10 VOLTAGE OF BATTERY ON CLOSED CIRCUIT

[9]

TABLE II

RESISTIVITY OF VARIOUS METAL CONDUCTORS AT 20° C.

MATERIAL	RESISTIVITY (MICROHM-CM)
Silver	1.62
Copper (Annealed)	1.72
Copper (Hard-drawn)	1.77
Chromium	2.7
Aluminum	2.82
Tungsten (Annealed)	4.37
Molybdenum	5.7
Zinc	5.92
Brass	7.5
Cadmium	7.60
Iron (Pure)	10.
Platinum	10.0
Nickel (Commercial)	11.0
Phosphor Bronze	11.5
Tin	11.5
Tantalum	15.5
Lead	22.0
German Silver	33.8
Mercury	95.8

(measured while the circuit is closed). We learn, therefore, that **the sum of the potential differences measured across all parts of the circuit, beginning at the positive pole of the battery and returning to the negative, is equal to the voltage of the battery,** or we might say, the applied voltage distributes itself proportionately throughout the closed circuit. If in Figure 2-11 the value of the resistance from a to d and the voltage of the electromotive force are known, it is possible to calculate the resistance of that part of the circuit between a and b from the voltmeter reading.

Example: The total resistance of the circuit shown by Figure 2-11 is 15 ohms, the voltage of the electromotive force on closed circuit is 10 volts, the potential drop across R_1 is 3 volts; what is the resistance of R_1?

Solution: For entire circuit:

$$I = \frac{E}{R} = \frac{10}{15} = .67 \text{ ampere.}$$

For the part of the circuit in question—

$$E = 3 \text{ volts}$$

Therefore,

$$I = .67 \text{ ampere}$$

$$E = 3 \text{ volts}$$

$$R_1 = \frac{E}{I} = \frac{3}{.67} = 4.5 \text{ ohms, ans.}$$

2.11 Internal Resistance

If a dry cell, as shown in Figure 2-9, is placed in a closed circuit like that of Figure 2-1 and its voltage again measured with a voltmeter, a reading will be obtained which will be somewhat less than the reading on open circuit. This means that the electromotive force of the dry cell depends to some extent upon the value of the current it is furnishing. As the current is increased the electromotive force is decreased. This is due to a potential drop within the cell itself, which is merely a drop across a resistance, excepting that in this case the resistance is inside the dry cell. Any electrical current leaving the positive pole of the dry cell and returning to the negative pole from the external circuit must likewise flow from the negative to the positive through the chemicals in the dry cell. These chemicals have a definite resistance called the **internal resistance.** In our consideration of the simple circuit, therefore, we **must either use the electromotive force measured on closed circuit or recognize that the open circuit electromotive force is acting through a resistance additional to that of the external circuit.** The absolute convention for this source of
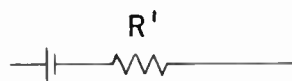
R'

FIGURE 2-12

FIG. 2-11 VOLTAGE DROP ACROSS ONE RESISTANCE OF CLOSED CIRCUIT

[10]

TABLE III

ELECTRICAL PROPERTIES OF COPPER CONDUCTORS STANDARDIZED BY LONG LINES DEPARTMENT

CONDUCTORS	NO.	SIZE		WEIGHT	RESISTANCE	
		GAGE	DIAMETER IN INCHES	LBS. PER WIRE MILE	*OHMS PER LOOP MILE	OHMS PER 1000 FEET (SINGLEWIRE)
Open Wire	8	B. W. G.	.165	435	4.10	.389
	10	N. B. S. G.	.128	264	6.82	.645
	12	N. B. S. G.	.104	174	10.3	.975
Cable	10	A. W. G.	.102	168	10.1	.955
(side circuits of standard	13	A. W. G.	.072	82.6	20.5	1.99
quadded cable)	16	A. W. G.	.051	41.2	40.2	3.81
	19	A. W. G.	.036	20.5	84.1	7.95

* The resistance values for open wire circuits are for 20° C or 68° F. Add 2/10 of 1% per degree Fahrenheit for higher temperatures. The resistance values for cable circuits are for 12.80° C or 55° F.

NOTE: A. W. G. is American Wire Gage and is same as B. & S. which is Brown and Sharpe Gage. B. W. G. is Birmingham Wire Gage and N. B. S. G. is New British Standard Gage.

electromotive force would be that shown by Figure 2-12, which represents the open circuit voltage plus a series resistor equal to the internal resistance of the cell.

The ordinary dry cell has an internal resistance averaging about one ohm, but this greatly increases with the aging of the cell. In the telephone central office storage batteries are used almost exclusively because the internal resistance is negligible for most direct current considerations.

2.12 Electric Power

In the simple circuits we have thus far considered we have only dealt with resistance, electromotive force, and electric current, but each of these circuits is actually converting energy from chemical to heat or some other form. They, therefore, have a definite power consumption or represent a definite transfer of power to some external device. **In the electric circuit if we multiply the electromotive force in volts by the current in amperes we have an expression for the power in watts.** The watt may, therefore, be defined as the power expended in a circuit having an electromotive force of one volt and a current of one ampere.

The following equation should be considered second only to Ohm's Law in importance:

$$P = EI \qquad (2:4)$$

A somewhat more convenient form for determining the power expended in any given resistance is

$$P = I^2R \qquad (2:5)$$

This latter equation is apparent from Ohm's Law, which states that $E = IR$ and we may, therefore, substitute IR for E in equation (2·4), which gives us I^2R.

Example: In Figure 2-10, what is the power expended in the resistor between terminals a and b if the potential difference is equal to 10 volts and the resistance is 5 ohms?

Solution: $P = EI$, and

$$I = \frac{E}{R} = \frac{10}{5} = 2,$$

then $\qquad P = EI = 10 \times 2 = 20$ watts, ans.

2.13 Properties of Electric Conductors

Dr. Ohm investigated the conducting properties of various kinds of metals and called those offering very high resistance to the flow of electricity "poor conductors" and those offering comparatively little resistance to the flow of electricity "good conductors". There is another classification for material having extremely high resistance, in fact so high as to give an open circuit for all practical purposes. These are called insulators.

[11]

Table II shows a few conductors in the order of conductivity. Those offering the least resistance are at the top of the list. Materials which are commonly used as insulators include: glass, mica, polystyrene, rubber, ebonite, cotton, silk, paper, bakelite, asbestos, porcelain and resins. There are many other good insulators but they are not all adaptable for use as such in practice.

In addition to the law showing the relation between electromotive force, current, and resistance, Ohm investigated the properties of conductors and established in addition to their relative values the following laws:

a. The resistance of any uniform conductor varies directly with its length.

b. The resistance of any uniform conductor varies inversely with its cross-sectional area.

Here we have the analogy to the water pipe previously mentioned but fortunately the electric conductors have more exact laws governing their electric resistances than water pipes have governing their resistance to the flow of water.

Copper is the most universally used conductor in electrical work. It offers very low resistance, does not deteriorate rapidly with age and has many mechanical advantages. There are several standard wire gages for designating the cross-sectional area or diameter of copper wire, and three apply to the standard conductors used by the Long Lines Department.

Table III shows the principal standard gages of wire used by the Long Lines Department and their resistance values.

Simple rules for remembering the approximate constants of the cable conductors are as follows:

a. Four sizes of cable conductors are standard for most uses in the Long Lines Department and all are A.W.G. (or B and S).

b. The largest size is #10 A.W.G. Add three gages for successive smaller sizes,—thus #10, #13, #16 and #19.

c. The diameter of #10 A.W.G. is slightly greater than one-tenth inch and its resistance is slightly greater than ten ohms per loop mile.

d. Smaller sizes double resistance by the addition of each three gages beginning with #10 as a base.

e. In cables, conductors are slightly longer than the cable lengths due to the spiraling effect. This will average about 5%.

f. Three sizes of conductors are standard for open wire; 104 (# 12 N.B.S.G.), 128 (# 10 N.B.S.G.) and 165 (#8 B.W.G.)

g. #10 is the nearest A. W. Gage to 104 (#12 N.B.S.G.) but is slightly smaller.

CHAPTER 3

THE SOLUTION OF D-C NETWORKS

3.1 Series Circuits

A simple circuit may contain any number of resistances. Figure 3-1 shows such a circuit with two resistors which, when connected as shown, are said to be **in series.** Figure 3-2 shows another circuit with the same resistors connected in **parallel.** Any number may be so connected in either case.
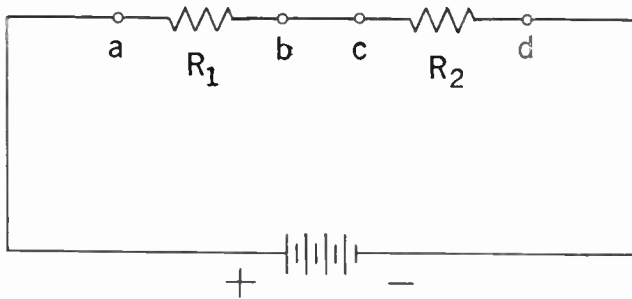
FIG. 3-1 SERIES CIRCUIT

The current from a battery in a parallel circuit will divide between the various resistance branches; but in a series circuit, as in the flow of water in a single pipe, it cannot divide and must be identical at every point. In other words, it must have an unchanged value in all parts of the circuit from the positive to the negative terminal of the battery. It follows then that **the total resistance of a series circuit is equal to the sum of all the individual resistances.** Thus, in Figure 3-1 the total resistance is the sum of the two resistances; namely, that connected between a and b or R_1, and that connected between c and d or R_2.

$$R = R_1 + R_2 \qquad (3:1)$$

3.2 Parallel Circuits

If we apply Ohm's Law to either of the two parallel resistors of Figure 3-2, we shall find that the current in either case must be equal to the potential measured across the particular resistor divided by its value in ohms; and for this particular circuit, the potential measured across either resistor is the emf of the battery. The battery is in reality supplying two currents, one through the resistor ab and the other through the resistor cd.

These two currents are united and flow together in the conductors connecting the poles of the battery with the junctions of the two resistors. For any circuit having two resistors connected in parallel, the current supplied to the combination must therefore be greater than the current supplied to either of the resistors. If we think of the combination of resistors in Figure 3-2 as equivalent to a single resistor that might be substituted in their stead, it is accordingly clear that the combined value in ohms of two resistors in parallel must be less than that of either resistor taken singly.

We may make calculations for determining the current in a parallel circuit such as is shown by Figure 3-2, but these are more complicated than for a simple series circuit having more than one resistance, such as is shown in Figure 3-1. The solution of a parallel circuit is accomplished with the aid of Kirchoff's Laws in addition to Ohm's Law.

3.3 Kirchoff's First Law

Kirchoff's First Law states that at any point in a circuit there is as much current flowing to the point as there is away from it. This applies regardless of the number of branches that may be connected to the point in question. The law can be interpreted by its application to point P in Figure 3-2. If I is the current being supplied by
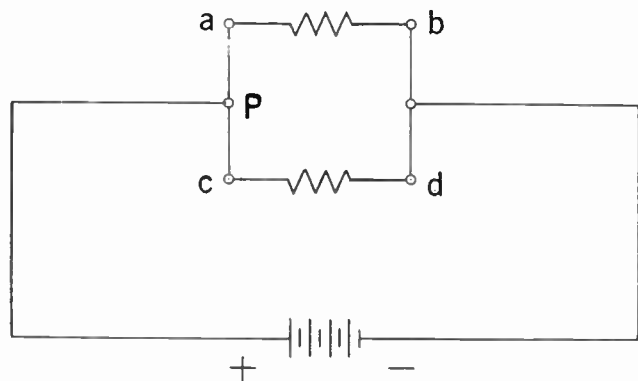
FIG. 3-2 TWO RESISTORS IN PARALLEL

the battery to the combination of the two resistors in parallel, and I_1 and I_2 are the respective currents through the two parallel resistors, then

$$I = I_1 + I_2 \qquad (3:2)$$

If we apply Ohm's Law to the entire circuit and let R represent the value of the combined resistances in parallel, we have

$$R = \frac{E}{I} \text{ or } R = \frac{E}{I_1 + I_2}$$

But

$$I_1 = \frac{E}{R_1} \text{ and } I_2 = \frac{E}{R_2}$$

Therefore,

$$R = \frac{E}{\dfrac{E}{R_1} + \dfrac{E}{R_2}}$$

In this latter equation, the E's can be cancelled and the equation written—

$$R = \frac{1}{\dfrac{1}{R_1} + \dfrac{1}{R_2}} \qquad (3:3)$$

and if we simplify this compound fraction by simple algebra—

$$R = \frac{R_1 R_2}{R_1 + R_2} \qquad (3:4)$$

This gives an equation for calculating the combined value of two parallel resistors. Expressed in words it may be stated as follows: **To obtain the combined resistance of any two resistors in parallel, divide their product by their sum.**

Example: What is the combined resistance of the inductive and non-inductive windings of a type-B relay used in a local A-board cord circuit if the inductive winding measures 16.4 ohms and the non-inductive winding measures 22 ohms?

Solution:

$$R = \frac{R_1 R_2}{R_1 + R_2} = \frac{16.4 \times 22}{16.4 + 22} = 9.4 \text{ ohms, ans.}$$

Figure 3-3 shows a circuit having three resistors in parallel. An equation similar to (3:3) can be worked out for combinations of this kind, or calculations can be made to obtain the combined resistance of *ab* and *cd* and this value then combined with the value of *ef*. But for problems involving more than two resistors in parallel, it is usually simpler to use the conductance method.
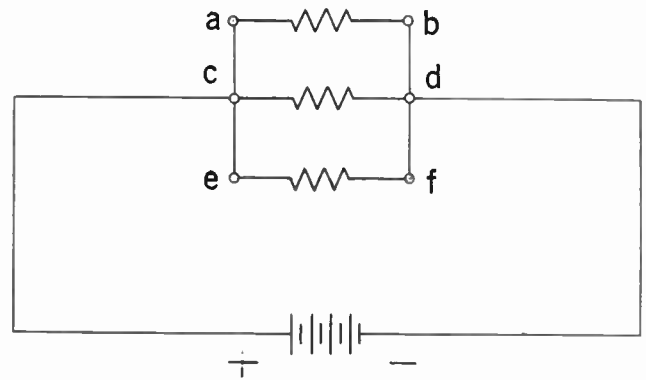
FIG. 3-3 THREE RESISTORS IN PARALLEL

3.4 Conductance

Conductance is defined as the reciprocal of resistance. It is expressed by the symbol G, and for any single resistance—

$$G = \frac{1}{R} \qquad (3:5)$$

For a combination of resistors in parallel, such as is shown by Figure 3-3, the conductance of the combination is equal to the sum of the individual conductances, or

$$G = G_1 + G_2 + G_3 \qquad (3:6)$$

In a circuit having a number of resistors in parallel, it is often of advantage to solve for the total conductance of the circuit and then find its total resistance by taking the reciprocal of the total conductance.

Example: If a B-3 relay has an inductive winding of 16.4 ohms, a non-inductive winding of 31 ohms, and these are shunted by an 18-U resistance (of 100 ohms), what is the resistance of the combination?

Solution:

$$G_1 = \frac{1}{R_1} = \frac{1}{16.4} = .061$$

$$G_2 = \frac{1}{R_2} = \frac{1}{31} = .032$$

$$G_3 = \frac{1}{R_3} = \frac{1}{100} = .010$$

$$G = G_1 + G_2 + G_3$$
$$= .061 + .032 + .010 = .103$$

$$R = \frac{1}{G} = \frac{1}{.103} = 9.7 \text{ ohms, ans.}$$

| 14 |
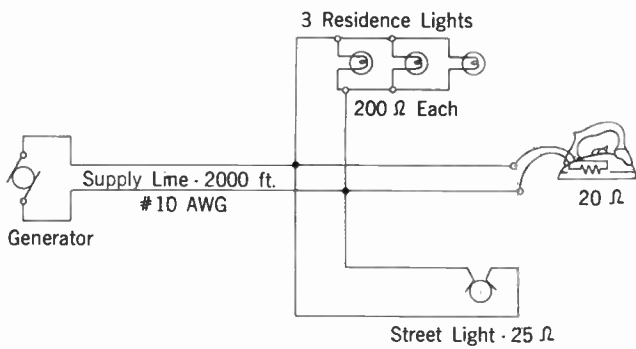
FIG. 3-4 SMALL ELECTRIC POWER SYSTEM

3.5 Direct-Current Networks

Several resistors may be connected in such a manner as to form very complicated networks. In practice many circuits are of this type. For example, Figure 3-4 illustrates a 110-volt power distribution line supplying a residence and a street light. We may represent the electrical characteristics of such a circuit by the network shown by Figure 3-5, and can further simplify this network as shown by Figure 3-6. Power supply systems are usually complicated networks of this sort.

In the same way, many telephone circuits may be analyzed by drawing their equivalent network diagrams. Figure 3-7 represents an A-board local cord circuit connected to a local switching trunk having $\frac{1}{2}$ mile of 19-gage cable. The equivalent network is shown by Figure 3-8.

In the solution of d-c networks, it is usually desired to know the current in the various branches, having given the resistance values of each individual branch and the voltage of the source or sources of emf.

Example: What is the value of the current through each winding of the B-59 relay in Figure 3-7?
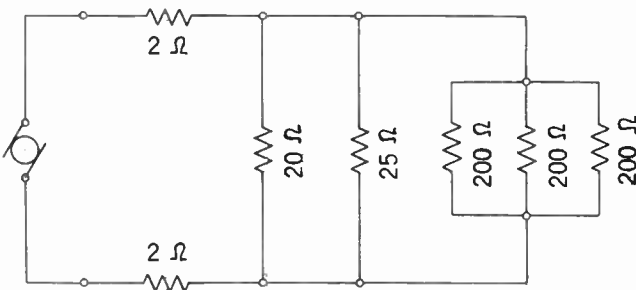
FIG. 3-5 CONVENTION FOR CIRCUIT OF FIG. 3-4

Solution: We must first find the total current through both windings and have:

$$I = \frac{E}{R} \text{ where } E \text{ is 24 volts and}$$

$$R = 139.4 + \frac{R_1 R_2}{R_1 + R_2}$$

$$= 139.4 + \frac{27 \times 12000}{27 + 12000}$$

$$= 139.4 + 26.9 = 166.3 \text{ ohms}$$

Then $\quad I = \frac{24}{166.3} = .14432$ ampere.

But the potential drop V across the two windings is equal to the current times the combined resistance of the two windings, or

$$V = I \times \frac{R_1 R_2}{R_1 + R_2}$$

$$= .144 \times 26.9$$

$$= 3.88 \text{ volts}$$
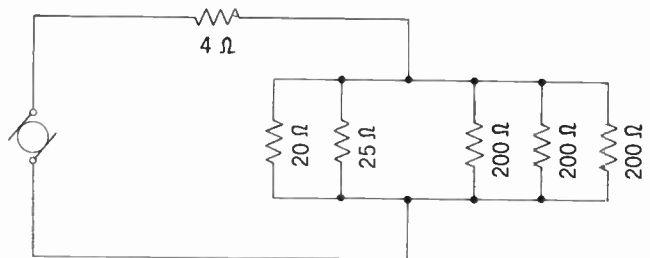
FIGURE 3-6

Then, applying Ohm's Law to each winding independently, we have—

$$I_1 = \frac{V}{R_1} = \frac{3.88}{27}$$

$$= .144 \text{ ampere, ans.}$$

and

$$I_2 = \frac{V}{R_2} = \frac{3.88}{12000}$$

$$= .00032 \text{ ampere, ans.}$$

3.6 Kirchoff's Second Law

When current flows through a resistor there is always a difference in potential between the ends of the resistor, the value of which depends upon the current flowing and the value of the resist-
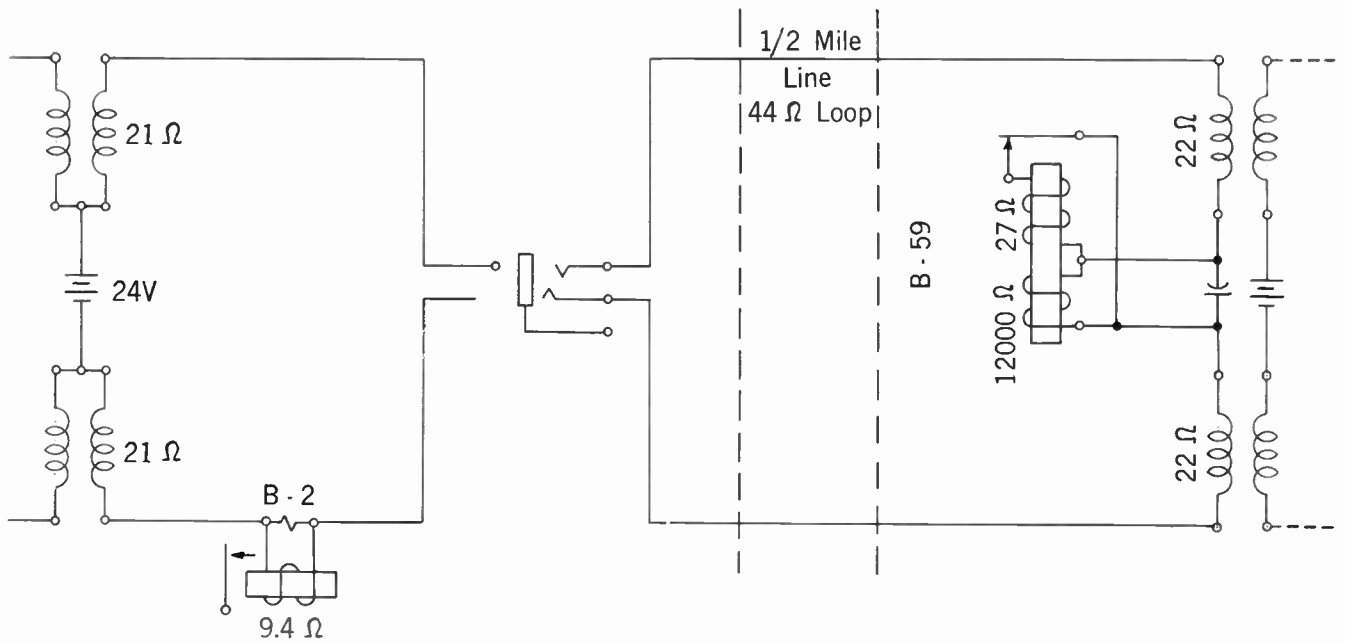
FIG. 3-7 CORD CIRCUIT AND SWITCHING TRUNK OF TELEPHONE CONNECTION

ance. This difference in potential is commonly called the **IR drop** since it is equal to the product of the current and the resistance. This *IR* drop acts in the opposite direction to, or opposes, the emf which drives the current through the resistor.

In a closed circuit, such as is shown in Figure 3-9, the sum of the *IR* drops across the three resistors must be equal to the impressed emf. Thus if the drop across the resistor R_1, as measured by the voltmeter, is represented by V_1 and those across R_2 and R_3 by V_2 and V_3 respectively, we may write the following equation—

$$E = V_1 + V_2 + V_3 \qquad (3:7)$$

This fact is known as Kirchoff's Second Law, which states that for any closed circuit or any closed portion of a complicated circuit, the algebraic sum of the emf's and the potential drops is equal to zero.

In the case of Figure 3-9, Kirchoff's Second Law may be written as follows:

$$E - IR_1 - IR_2 - IR_3 = 0 \qquad (3:8)$$

In solving any network problem, the first thing to do is to draw a good diagram. When the problem is to be solved by Kirchoff's Laws, the next step is to assign letters to all the unknowns in the circuit and to put arrows on the circuit diagram to indicate the **assumed directions** of current flow. If Kirchoff's First Law is applied at the junction points of a network, the number of unknowns may be kept down. Thus, if three wires meet at a point, and I_1 and I_2 have already been assigned to the currents in two of them, the third current may be designated as their sum or difference, depending upon the assumed direction of current flow. That is, instead of using a third unknown I_3, we will have $(I_1 + I_2)$ or $(I_1 - I_2)$. This will eliminate one equation. However, at least as many equations as there are unknowns must be written.

In the practical application of Kirchoff's Laws, the correct use of algebraic signs is fundamentally important. When one sign has been given to the electromotive force in the direction of the current flow, the opposite sign must be given to the *IR* drops. In other words, when going through a

FIG. 3-8 SIMPLIFIED CONVENTION FOR CIRCUIT OF FIG. 3-7

resistance in the same direction as the current flow, there is a drop in voltage and this voltage should be preceded by a minus sign. Conversely, when going through a resistance in the direction opposite to the current flow, there is a rise in voltage which should be preceded by a plus sign. We may for convenience accept the clockwise direction as positive, or **accept as positive all emf's which tend to make a current flow in a clockwise direction, and as negative all potential drops due to this flow of current as well as any emf's in the circuit tending to make current flow in the opposite direction.** It is immaterial whether the directions of current flow assumed are actually correct, as long as they are consistent throughout the network. The signs of the answers will show whether or not the assumed directions are correct. When the value of a current found by solving the equations is preceded by a minus sign, it merely means that the actual direction of flow is opposite to the direction which was assumed.

FIGURE 3-9

Example: Find the current values in each branch of Figure 3-10, if the resistance of $R_1 = 5$ ohms, $R_2 = 10$ ohms, $R_3 = 15$ ohms, and $R_4 = 20$ ohms, and the voltage $E = 24$ volts.

Solution: We may first assume that the direction of current flow is clockwise through both branches of the network. Applying Kirchoff's First Law at the point b, we know that the current flowing through R_3 plus the current flowing through R_2 and R_4 equals the total current flowing through R_1. Therefore I_1, the current through R_1, is equal to $I_2 + I_3$.

Considering first only the one closed loop of the circuit of Figure 3-10 that is shown by

[17]

FIG. 3-10 SERIES-PARALLEL CIRCUIT

Figure 3-11, we may write in accordance with Kirchoff's Second Law.

$$E - R_1 (I_2 + I_3) - R_3 I_3 = 0 \qquad (a)$$

and for the closed loop shown by Figure 3-12

$$E - R_1 (I_2 + I_3) - R_2 I_2 - R_4 I_2 = 0 \qquad (b)$$

We thus have two independent equations containing two quantities which are unknown, namely, I_2 and I_3. Substituting the known values of E, R_1, R_2, R_3, and R_4, these equations may be written as follows:

$$24 - 5 (I_2 + I_3) - 15 I_3 = 0 \qquad (a)$$

$$24 - 5 (I_2 + I_3) - 10 I_2 - 20 I_2 = 0 \qquad (b)$$

Simplifying, these equations become—

$$24 - 5 I_2 - 20 I_3 = 0 \qquad (a)$$

$$24 - 35 I_2 - 5 I_3 = 0 \qquad (b)$$

Multiplying equation (a) by seven, and subtracting equation (b) from it, we have—

$$168 - 35 I_2 - 140 I_3 = 0 \qquad (a)$$

$$24 - 35 I_2 - 5 I_3 = 0 \qquad (b)$$

$$\overline{\quad 144 \qquad\qquad - 135 I_3 = 0 \quad}$$

or

$$I_3 = \frac{144}{135} = 1.07 \text{ amperes, ans.}$$

FIGURE 3-11

a I_1 b I_2 c

CHAPTER 4

MAGNETS AND MAGNETIC CIRCUITS

4.1 Nature of Magnetism

The early Greeks were familiar with a natural stone that would attract bits of iron. It was a form of iron ore, now known as magnetite, and the power of attraction possessed by it was called **magnetism**. It was also discovered at an early date that a bar or needle-shaped piece of this ore, when suspended so as to turn freely about a central axis, would turn so that one end always pointed to the north. For centuries, before it was discovered that there was any relationship between magnetism and electricity, this unique property was put to practical use in the mariner's compass —as it still is.

Investigation through the years showed that this magnetic property could be induced by one means or another in iron and steel, and to a lesser degree in nickel and cobalt, as well as in alloys of these metals. Not until early in the nineteenth century was it learned that the magnetic property could be artificially given to these metals by means of an electric current.

Magnets, as we know them today, are classed as **permanent magnets** and **electromagnets**. A hard steel bar when magnetized becomes a **permanent** magnet because it tends to retain its magnetism under normal conditions for a long period unless subjected to heat or jarring. Soft iron tends to become easily magnetized when subjected to a magnetizing influence, but does not retain an appreciable part of the magnetism thus imparted

to it. Consequently, permanent magnets are of steel or of such an alloy as cobalt-steel or **remalloy** (iron-cobalt-molybdenum), and cores for electromagnets are ordinarily made of **soft iron** or of iron alloys such as **permalloy** (iron-nickel), **supermalloy** (molybdenum-permalloy), **perminvar** (cobalt-iron-nickel), or **permendur** (iron-cobalt).

4.2 Permanent Magnets

Figure 4-1 represents a rectangular steel bar magnet which will attract bits of iron brought near to either end, and will exert a force of either repulsion or attraction upon other magnets in its

FIGURE 4-2

vicinity. The influence of a magnet may be detected in the space surrounding the magnet in various ways, and is found to vary inversely as the square of the distance from the magnet. To account for this phenomenon, the magnet is said to have a magnetic field, which is represented by the curved lines in Figure 4-1. These curved lines are merely a convention for illustrating the effect of the magnet. They are commonly known as **lines of magnetic induction.** All the lines as a group are referred to as the flux, and designated by the symbol ϕ. The flux per unit area is known as the **flux density** and is designated by B.

The lines of magnetic induction are thought of

Fig. 4-1   Magnetic Field Around Bar Magnet

[20]
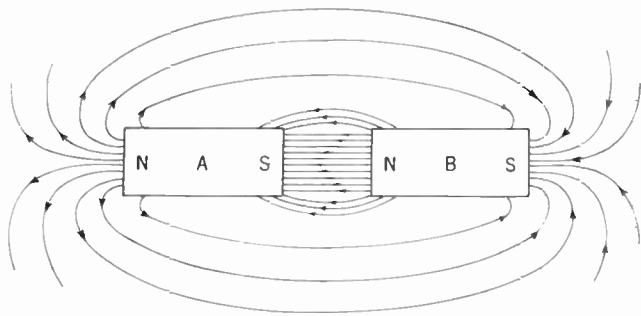
as passing through a magnet from the south to the north pole, leaving the magnet at its north pole and reentering the magnet at its south pole. This is the significance of the arrows shown on the lines in Figure 4-1.

Lines of magnetic induction are always closed loops. A vivid graphical demonstration, not only of the presence of the magnetic field but of the arrangement of the lines of magnetic induction, may be had by sprinkling iron filings upon a glass plate placed above a magnet. Figure 4-2 shows how the filings arrange themselves under such a condition.

If a second magnet is placed at the end of the bar magnet shown in Figure 4-1, the magnetic field will become either like that shown in Figure 4-3 or that shown in Figure 4-4. In the first case the two magnets will attract each other. In the second case they will repel each other. If they should attract and establish a combined magnetic field such as that shown by Figure 4-3, merely changing ends of one magnet will give the effect in Figure 4-4. We then learn from the action of one magnet toward another that the two ends of any magnet are unlike. These two ends are called the **poles** and for convenience, the pole having one influence is called the **north pole** and that having the opposite influence is called the **south pole**. The distinction comes from the earth, which is itself a magnet. When a bar magnet is suspended so as to swing freely, that pole which tends to point toward the north is called the north-seeking or north pole; the other is called the south pole. The needle of the surveyor's compass is an application of a bar magnet free to swing on its pivot, and its north pole will point to the earth's magnetic pole located near the geographical north pole. (However, since the earth is itself a magnet, it may be noted that with this conventional definition, the pole nearest the geographical north is

the earth's south magnetic pole inasmuch as it attracts unlike or north-seeking poles of suspended magnets.)

If the strength of the magnet in Figure 4-1 is doubled, the magnetic field will be strengthened in proportion, and may be represented by a more congested arrangement of lines of magnetic induction. The force that will be exerted upon a pole of another magnet located at any point in the magnetic field will depend upon the **intensity** of the field at that point. This field intensity is represented by the symbol H.

We have said in a preceding paragraph that the flux density B is the number of lines of magnetic induction passing through a unit area. By definition, unit flux density is one line of magnetic induction per square centimeter. We have also said that lines of magnetic induction are merely conventions for illustrating the effect of a magnetic field. Such a line may therefore be defined arbitrarily. In practice, it is usually defined as that magnetic induction per square centimeter in air, which exists in a magnetic field having unit intensity. Thus in air the field intensity H and the flux density B have the same numerical value.

In Figure 4-1 we see that the magnetic field has greatest intensity nearest the poles. If we wish to create a field of greater intensity, we can accomplish it by bending the magnet into the form of a horseshoe like that shown in Figure 4-5. Here each line emerging from the north pole returns to the south pole of the magnet through a much shorter distance than that represented by any one of the curved loops in Figure 4-1. The strength of the field between the two poles of a horseshoe magnet is more intense than that of a straight magnet of equal strength. Thus, we not only shorten each line represented by a closed loop but, in so doing, create more lines. This gives us an analogy to the electric circuit, which we may call a **magnetic circuit**. In the electric circuit, if
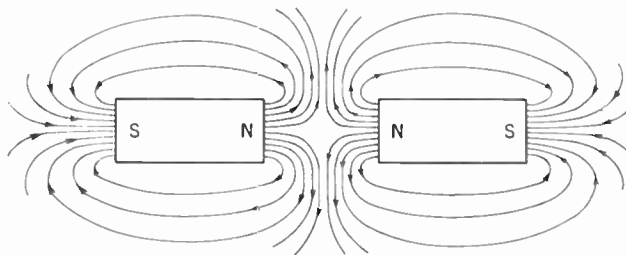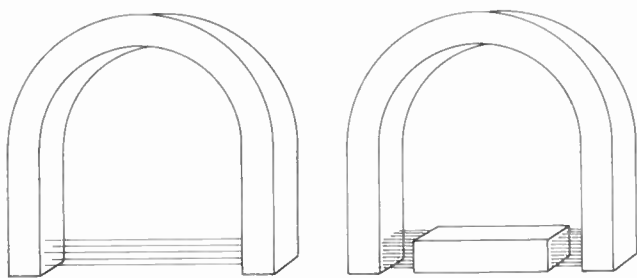
FIGURE 4-5

we have a conductor connected between the positive and negative poles of a battery and decrease the resistance by decreasing its length, we increase the current strength. In the case of the magnet, if we decrease the lengths of the paths from the north to the south pole by bending the magnet into the form of a horseshoe, we increase the number of lines of magnetic induction.

Again, if we insert between the poles of the horseshoe magnet in the space now filled with air, a piece of soft iron or other magnetic material, we greatly increase the number of lines of magnetic induction existing in the circuit formed by the magnet itself and the soft iron used for closing this circuit between the north and south poles. This is analogous to decreasing the resistance of an electric circuit by substituting a conductor of lower resistance for one of higher resistance.

4.3 The Magnetic Circuit

As electric current is caused to flow in an electric circuit, so magnetic flux can be established in a magnetic circuit. Magnetic flux ϕ, or the total number of lines of induction existing in the circuit, then, is in some respects analogous to electric current.

The flux density B, or the number of lines of induction per unit area, may be written—

$$B = \frac{\phi}{A} \qquad (4:1)$$

where A is the area taken at right angles to the direction of the flux and ϕ is the flux through and normal to this area.

Since the lines of magnetic induction are increased by the insertion of some material other than air in the magnetic field, it follows that the flux density depends upon the materials of the completed magnetic circuit and the strength of the magnet. This corresponds to an electric circuit wherein the current density in any given

cross-section of conductor depends upon the resistance of the closed circuit and the electromotive force applied. There is, therefore, a property of a magnetic circuit which is analogous to the resistance of an electric circuit. This property is called **reluctance.** Likewise, there is a property of the magnet which is analogous to the electromotive force of a battery. This is called the **magnetomotive force.** For the complete magnetic circuit, we may apply an equation identical in form to Ohm's Law which, in words, may be stated—**the flux for any given magnetic circuit is equal to the magnetomotive force of the magnet divided by the reluctance of the circuit.** Expressed mathematically, this may be written—

$$\phi = \frac{M}{R} \qquad (4:2)$$

where the symbol for flux is ϕ, for magnetomotive force, M, and for reluctance, R. This may be compared to Ohm's Law as expressed by equation (2:1).

$$I = \frac{E}{R}$$

While we see that in many respects the magnetic circuit is analogous to the electric circuit, it is well to remember that the analogy is not complete, since there are other respects in which the two circuits differ. The two more important of these to bear in mind are as follows:

(a) A magnetic circuit can never be entirely opened; a magnetic field must exist at all times in the vicinity of a magnet. For this reason the magnetic circuit would be more nearly analogous to the
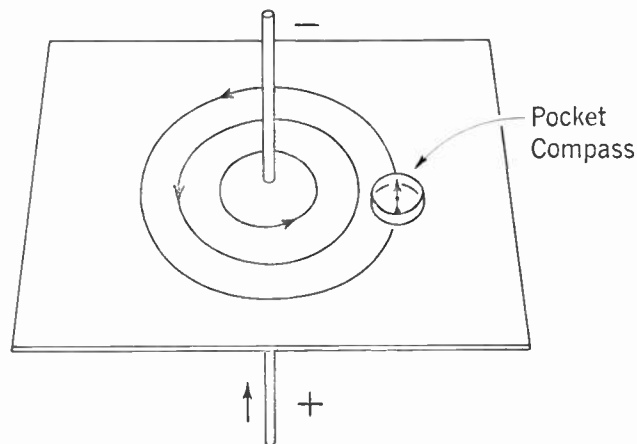
FIG. 4-6 MAGNETIC FIELD AROUND CURRENT-CARRYING STRAIGHT CONDUCTOR

electric circuit submerged in water. When the continuity of the metal conductors forming such an electric circuit is broken, the circuit will be completed through the liquid across its gap. Though the current strength might be decreased in this way, the circuit could never be entirely opened; neither would the current be limited to the submerged metal conductors. There would be other flow surrounding the conductors but not of such great intensity as in the metal conductors.

(b) Flux is not strictly analogous to current since current is rate of flow of electricity while the nature of flux is more nearly a state or condition of the medium in which it is established.

4.4 Electromagnets

If a conductor carrying an electric current pierces a cardboard as shown in Figure 4-6, there may be detected on the plane of the cardboard a magnetic field with lines of magnetic induction encircling the conductor. To illustrate further, if iron filings are sprinkled on the cardboard, they will form visible concentric circles as shown by Figure 4-7. Through such observations as these, we learn that **wherever an electric current is flowing there is also present a magnetic field, and the loops formed by the encircling lines of magnetic induction are always in a plane perpendicular to the conductor.**

If in either Figure 4-6 or 4-7 a compass is placed near the conductor, the needle will align itself tangent to some one of the many concentric
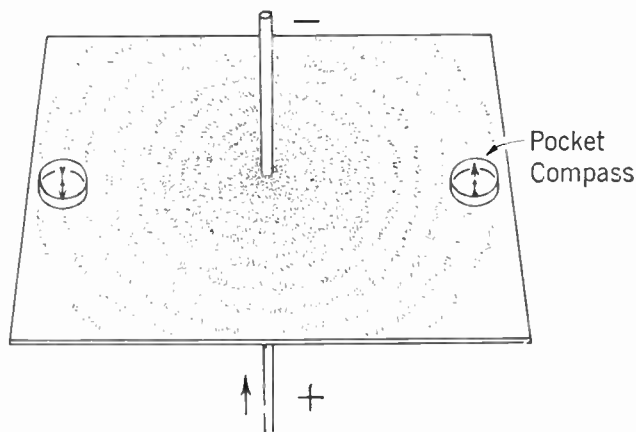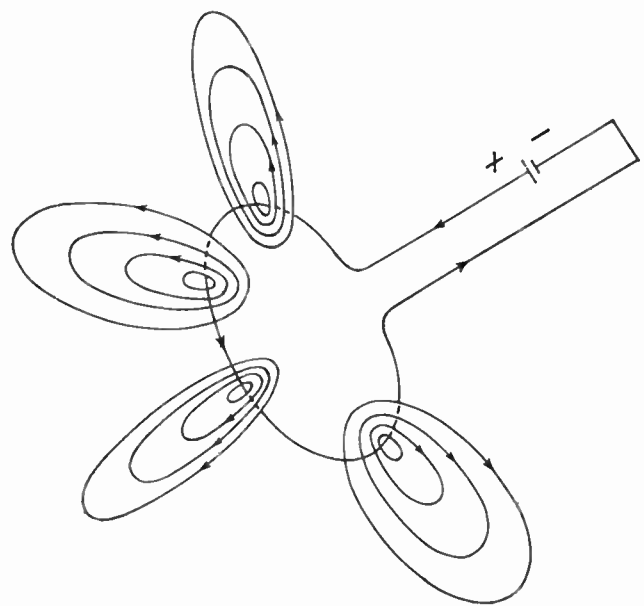
FIG. 4-8 MAGNETIC FIELD AROUND CURRENT-CARRYING LOOP

circles. If the compass is moved slowly around the wire, the needle will revolve on its pivot and maintain its tangential relationship. It will also be found that the direction of the lines with respect to the direction of current flow is that represented by the arrows in Figure 4-6.

Though this magnetic effect is a positive one, under the conditions shown in the Figures and even with a very strong current in the conductor, the magnetic field represented by the concentric circles is relatively weak. But if the electric conductor is made to form a loop, the groups of lines forming concentric circles for every unit of the conductor's length can be imagined as arranging themselves as shown in Figure 4-8. The closed loops are no longer circular. They become more crowded in the space inside the loop of wire and less crowded in the space outside the loop of wire. Accordingly, the intensity of the magnetic field within the loop is increased. This may be more clearly seen by considering the single line which Figure 4-9 shows enclosed by imaginary boundaries both within and without the loop of wire. We may express the field intensity in terms of the cross-section of this imaginary bounding space. At the point p inside of the loop the intensity is such as to give one line for the area represented by the cross-section a, and at the point P outside of the loop the intensity is such as to give one line for an area represented by A.

Pocket Compass

FIGURE 4-7

[23]
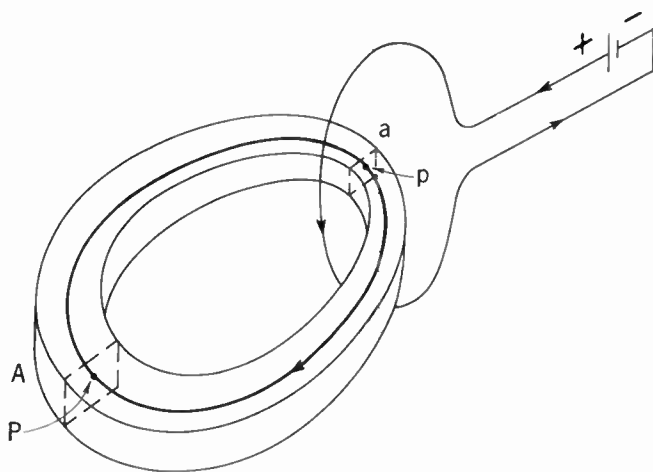
FIGURE 4-9

If, instead of having an electric circuit consisting of one loop of wire, we have a circuit consisting of several turns of wire such as the winding on the spool shown in Figure 4-10, the intensity of the field is multiplied by the number of turns of wire. Thus, the value of the field intensity at any point for two turns would be twice that for a single loop; for three turns, three times that for a single loop; and for *n* turns, *n* times that for a single loop, providing the turns are sufficiently close together so that flux leakage between successive turns is negligible.

Comparing Figure 4-10 with Figure 4-1, we find that the current in the coil of wire creates a magnetic field similar to that of the bar magnet. In Figure 4-6 the relationship between direction of current flow and direction of lines of induction was shown by arrows. We use this same relationship in Figure 4-8 and going one step further, we may determine the north and south poles of the magnet formed by the coil shown in Figure 4-10. A simple way to remember the relationship for any winding is illustrated by Figure 4-11. Here if we assume current flowing through a winding in the direction of "turn" for a right-hand screw, the lines leave the point of the screw, which is the north pole, and enter the slot, which is the south pole.

In Figure 4-5 the number of lines in the magnetic circuit established by the horseshoe magnet was greatly increased by the insertion of a piece of soft iron between the north and south poles. Likewise, if in Figure 4-10 the spool shown has a soft iron core, the number of lines will be greatly increased. Further, if the core of the winding is bent in the shape of a horseshoe as

shown in Figure 4-12, we have the customary electromagnet which is capable of exerting considerable force.

4.5 Relation Between Current and Field Intensity

If we increase the current strength in the winding shown by Figure 4-10, we will find that the intensity of the magnetic field is increased proportionately. Thus, the value of H, or the magnetic field intensity in air, is directly proportional to the current flowing in the winding. We may accordingly establish a definite relation between field intensity and electric current for any given set of conditions.

A winding such as that shown in Figure 4-10 is called a **solenoid**. If such a solenoid is very long as compared to its diameter, the field intensity in the air on the inside of the solenoid is directly proportional to the product of the number of turns and the current, and inversely proportional to the length of the solenoid. Expressed mathematically—

$$H = k \frac{NI}{l} \qquad (4:3)$$

where N is the total number of turns, I the current, l the length of the solenoid, and k is a constant depending on the units in which the other factors are expressed. In the rationalized mks (meter-kilogram-second) system of units, k is equal to 1, I is in amperes, l is in meters and H is then **ampere-turns per meter**.

The field intensity H may be thought of as the force tending to produce magnetic flux in each unit length of a magnetic circuit. It is the **magnetomotive force per unit length of circuit**. Its analogy in the electric circuit is the "distributed
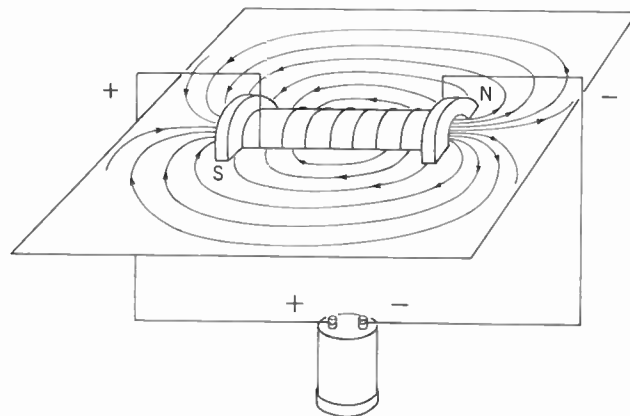
FIG. 4-10 MAGNETIC FIELD AROUND AIR-CORE SOLENOID

[24]

emf" per unit length of a uniform conductor, or that element of the electromotive force tending to force a current through each unit length of conductor. We may therefore express the total magnetomotive force of the solenoid in Figure 4-10 in ampere-turns, as the field intensity times the length of the solenoid—thus:

$$M = H \times l = kNI \qquad (4:4)$$

4.6 Flux Density, Field Intensity and Permeability

In discussing field intensity we have thus far considered it only in connection with magnetic circuits in air. However, we have seen that if iron is inserted in a solenoid such as that shown by Figure 4-10, the number of lines of induction will be greatly increased. This means that the flux density, or the number of lines per unit area of cross-section inside the solenoid, may be much greater than that set up in air by a field of the
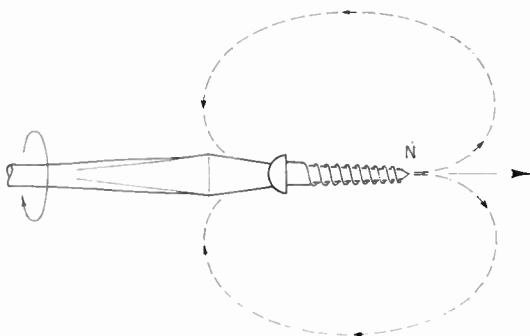
FIGURE 4-11

same intensity. In inserting the iron, we have greatly lowered the reluctance of the magnetic circuit. Because of the lowered reluctance, the magnetomotive force has established a greatly increased flux. We find then that if iron is introduced into a magnetic circuit, the flux density will depend upon the intensity of the field in the air before the iron is inserted, and upon certain magnetic properties of the iron, or the adaptability of the iron for lowering the reluctance per unit of length.

As noted above, we may think of the field intensity H in air in the sense of a definite magnetizing force which will set up a greatly increased flux in any unit length of iron having a lower reluctance than air. Ordinarily we do not use the reluctance of iron per unit length but employ in-
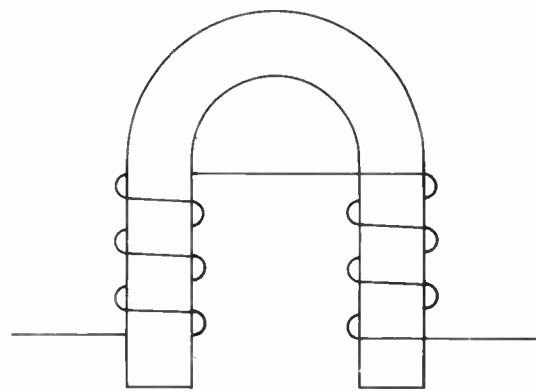
FIG. 4-12 HORSESHOE ELECTRO-MAGNET

stead a term which is inversely proportional to reluctance, or is analogous to conductivity in an electric circuit. This term is known as **permeability** and is represented by the Greek letter μ. It is the ratio of the magnetic conductivity of a substance to the magnetic conductivity of air Using this ratio, we may express the flux per unit cross-section in the form of an equation as follows:

$$\frac{\phi}{A} = H \times \mu \qquad (4:5)$$

or

$$B = H \times \mu \qquad (4:6)$$

where B is the conventional symbol for flux density.

This equation may also be written in other forms:

$$\mu = \frac{B}{H} \qquad (4:7)$$

or

$$H = \frac{B}{\mu} \qquad (4:8)$$
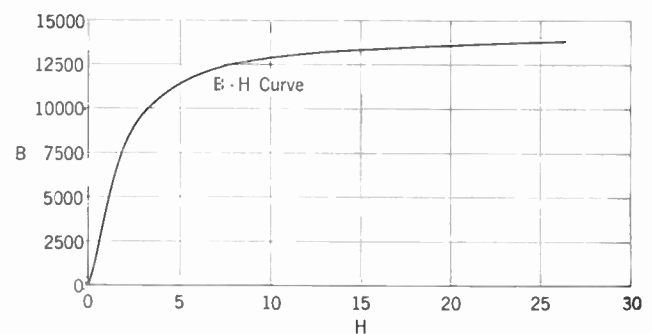
FIG. 4-13 B-H CURVE FOR IRON
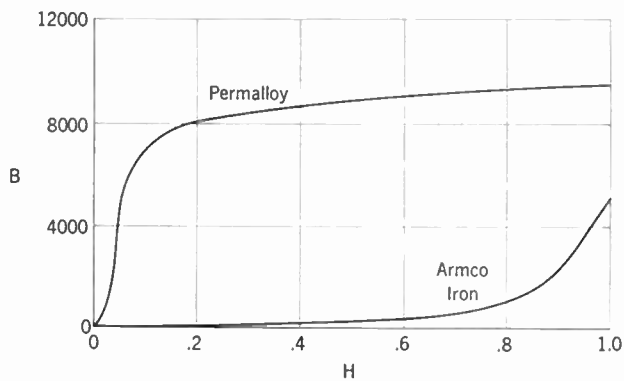
FIG. 4-14 B-H CURVE FOR PERMALLOY

4.7 Magnetic Properties of Iron

Permeability has been compared to electric conductivity. There is one distinction, however, which is most essential. The stability of iron under various degrees of magnetization is not equal to that of the ordinary metallic electric conductor. In the electric circuit, the resistance or conductivity remains very nearly fixed for any degree of current strength, unless there is some change in temperature. While the same may be said of the magnetic circuit in air, in iron the condition is different. As the number of lines of induction are increased (or the flux density is increased), the permeability of the iron is changed, and any further increase in the magnetizing force (or field intensity) may not mean a proportional increase in the flux density. In simpler terms, that property of the iron which enables it to establish more lines of induction depends entirely upon the number of lines that it already has. After a certain number per unit area of cross-section, or a certain flux density, the iron becomes less effective and regardless of any further increase in

FIG. 4-15 HYSTERESIS LOOPS FOR SOFT IRON AND STEEL

field intensity, the flux density may have already become so great that additional lines cannot be established any more readily than if the core were of air. This condition is called the "saturation point" of the iron.

4.8 B-H Curves

What is said in the preceding Article with respect to the magnetic properties of iron applies likewise to the other magnetic materials, and also to all of the magnetic alloys. Table II shows the resistance of electric conductors compared with copper. A similar table could be compiled for electric conductivity by taking the reciprocal of the resistance values shown. Such a table would

FIG. 4-16 HYSTERESIS LOOPS FOR MAGNETIC IRON AND PERMALLOY

be analogous to a magnetic table for permeability; but to give accurately the permeability for any magnetic material, it is necessary to show a complete curve rather than a single tabulated value. Such a curve is illustrated by Figure 4-13 which is taken for a magnetic iron used by the Western Electric Company in the manufacture of certain relays and other telephone apparatus. This curve was determined after the iron had been annealed for three hours at a temperature of 900° C. Every magnetic material has some such curve. A magnetization curve will ordinarily depend upon many things, such as—

    (a) Whether cast iron, wrought iron, steel or an alloy of these with other metals.

    (b) Degree of purity.

(c) Heat treatment used in preparing the metal.

(d) Previous magnetic history; that is, whether or not it has been subject to a high degree of magnetization in the past.

At low values of field intensity (H below 1.0) the magnetic material, permalloy, which is an alloy of nickel and iron (plus a small amount of chromium or molybdenum in certain cases) has a very much higher permeability than iron, making it extremely useful in communication work where low values of field intensity are common. Figure 4-14 gives B-H curves for a standard permalloy and a standard iron for low values of H; it will be noted that the magnetic flux for a given magnetizing force is very much greater in the permalloy than in the iron over the range covered.

4.9 Hysteresis

If a piece of iron is subjected to an increasing magnetizing force until the saturation point is reached and then the magnetizing force is decreased to zero and established in the opposite direction until the saturation point is again reached, and if the magnetizing force is again decreased to zero and again increased until the cycle is completed, the relations between flux density and field intensity for all parts of the cycle may be represented by a curve such as one of those shown by Figure 4-15. This is called the **hysteresis loop.** Here it is seen that after the iron has once reached the saturation point, it does not return to its original magnetic condition no matter to what magnetizing forces it may be subjected. For example, an inspection of the hysteresis loop shows that iron will retain a certain degree of magnetization after the magnetizing influence has been reduced to zero. This is particularly true of hard steel and is the reason that all permanent magnets are made of hard steel or a material having similar characteristics. The two curves of Figure 4-15 illustrate the difference in the hysteresis loops of hard steel and soft iron. The fact that soft iron has a narrow hysteresis loop makes it adaptable for the cores of electromagnets. We may note here also that the hysteresis loop for permalloy is very much narrower than that for soft iron at low values of magnetizing force. This is illustrated in Figure 4-16 where the hysteresis loop for permalloy and iron are compared.

CHAPTER 5

ELECTRICAL MEASUREMENTS IN DIRECT-CURRENT CIRCUITS

5.1 Measuring Instruments

We have been discussing such electrical quantities as the volt, the ampere, the ohm and the watt, but little has been said about the electrical instruments that are used to measure these quantities. Basic instruments of this kind include the galvanometer, the voltmeter, the ammeter, the Wheatstone bridge (including a galvanometer) the megger, and the wattmeter. At this stage of our study it is important that we learn the fundamental principles of these measuring instruments and the distinction between instruments designed for different purposes, but it is not important that we study long descriptions of their construction or those details of design pertaining only to their manufacture. They are ordinarily sealed at the factory and are seldom repaired by the field maintenance man. Let us, therefore, concern ourselves with the intelligent and skillful use of them and only with those principles of their operation that are essential to this.

The galvanometer may be considered the most elementary of electrical measuring instruments in that it is nothing more than a sensitive device for detecting electric (direct) currents. It is not designed to determine magnitudes of currents but merely their presence. Naturally its effectiveness in detecting currents of extremely small value depends upon its sensitiveness. While the galvanometer is the simplest of the group of instruments used in daily practice, it is one of the most delicate. It ordinarily consists of a coil of several turns of very fine wire suspended between the poles of a permanent horseshoe magnet and held in a neutral position by the torsion of fine suspension fibres, or other equally delicate means. The suspended coil carries a light needle which stands at the center of a fixed scale when the coil is in its neutral position with respect to the permanent magnet. A very small current through the suspended coil will set up a magnetic field that will tend to align itself with the field of the permanent magnet and thereby cause a deflection of the needle from its neutral position on the fixed scale.

In direct-current measurements, both ammeters and voltmeters are ordinarily galvanometer types of instruments. The ammeter is used for measuring the flow of current in a circuit. For this reason it must be inserted directly in the path of the current as was done with the water flowmeter shown in Figure 2-2 (Chapter 2). It can readily be seen that if the meter is to measure accurately the current flow in the circuit, it must not disturb to any appreciable extent the current or voltage relations in the circuit to be measured. To insure this, the resistance of the meter is made extremely low. Thus, when an ammeter is inserted in a circuit as in Figure 5-1, the current flow measured by the meter is equal to

$$\frac{E}{R_1 + R_2 + R_3 + r}$$

where r is the resistance of the meter. Since r is relatively insignificant, this value is not appreciably different from the current flow without the meter, which is equal to

$$\frac{E}{R_1 + R_2 + R_3}$$

For the same reason the voltage drops across the resistors, IR_1, IR_2 and IR_3, the sum of which must equal E, remain effectively undisturbed.

The voltmeter is used to measure voltage or potential drop. It is connected across a battery or a resistor across which exists a potential drop. This is also illustrated in Figure 5-1. Since the current or voltage relations in the circuit are also to remain undisturbed, the voltmeter must not
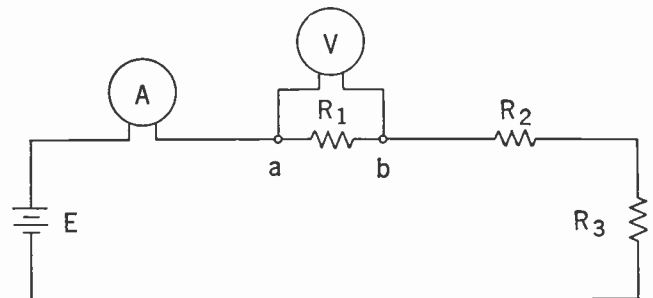
FIG. 5-1 VOLTMETER-AMMETER RESISTANCE MEASUREMENT

draw an appreciable current from the circuit across which it is connected. For this reason the voltmeter, unlike the ammeter, must have an extremely high resistance.

We have learned that the combined resistance of resistors in parallel is lower than that of the smallest resistor. It follows that a high resistor in parallel with a relatively small resistor will not appreciably change the total net resistance. This leads to a general rule for the use of voltmeters— **the resistance of a voltmeter must be very much higher than the resistance of the circuit across which the voltage drop is being measured.**
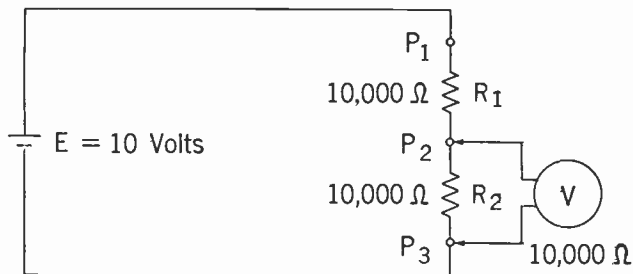
FIGURE 5-2

In order to show the error introduced by using a voltmeter whose resistance is comparable with that of the circuit element across which it is bridged, consider the circuit shown in Figure 5-2. If the voltmeter were not connected to the circuit, the current flowing in the circuit would be

$$I = \frac{E}{R_1 + R_2} = \frac{10}{10,000 + 10,000} = .0005 \text{ amps.}$$

The voltage drop either between points P_1 and P_2 or between P_2 and P_3 would be

$$IR_1 = IR_2 = .0005 \times 10,000 = 5 \text{ volts}$$

If, however, a voltmeter having a resistance of only 10,000 ohms is connected to P_2 and P_3 as shown, the total resistance between these points is the net parallel or combination resistance of the 10,000-ohm voltmeter and the 10,000-ohm resistor R_2; that is, 5,000 ohms. Then the total current in the circuit is

$$I = \frac{10}{10,000 + 5,000} = .000667 \text{ amps.}$$

Consequently, the voltage drop between P_2 and P_3 now is

$$IR_2 = .000667 \times 5,000 = 3.33 \text{ volts.}$$

Therefore, the voltage read by the voltmeter is only 3-1/3 volts instead of the 5 volts which would actually exist between P_2 and P_3 if the voltmeter were not used, or if a voltmeter were used whose resistance was much greater than the 10,000-ohm resistor across which it is bridged.

To avoid error, or when it is desired to measure the voltage across very high-resistance circuits, an "electron-tube voltmeter" is often employed. This is essentially a voltmeter that is associated with an electron-tube amplifier, the amplifier making it possible to obtain a full scale reading on the meter without drawing an appreciable amount of power from the circuit to which the voltmeter is connected. In other words, the input resistance of such a voltmeter may be practically infinite.

If a voltmeter measures at any given instant the emf across any direct current circuit (either branch or mains) and an ammeter at the same instant measures the current in the same circuit (either branch or mains), the product of the two readings is, from the equation $P = EI$, equal to the power in watts supplied to the circuit. Meters are designed with both ammeter and voltmeter terminals to read this product, or the power in watts directly. These are called wattmeters.

There are two remaining instruments in the basic group. These are the Wheatstone bridge and the megger. The megger is a combination of a magneto source of relatively high electromotive force and a sensitive meter, calibrated to read values of very high resistances connected across its terminals. The Wheatstone bridge is a network of resistors which is used in connection with a galvanometer for measuring an electrical resistance by an accurate comparison method. A more detailed discussion on the use of this device follows later in this Chapter.

In using any electrical instrument it must be remembered that Ohm's Law is never failing and that it applies to every circuit branch. Thus, in the case of the ammeter, the current that will flow through it will be very large if an appreciable potential is connected across its terminals without other resistance in the circuit. As an illustration, if an ammeter has an internal resistance of .005 ohms and an electromotive force of one volt is connected to its terminals, the current through it in accordance with Ohm's Law will be 200 amperes. This may be considerably in excess of the maximum current value for which the instrument is designed. It is well to remember, therefore, that the ammeter is an instrument that will cause

[29]
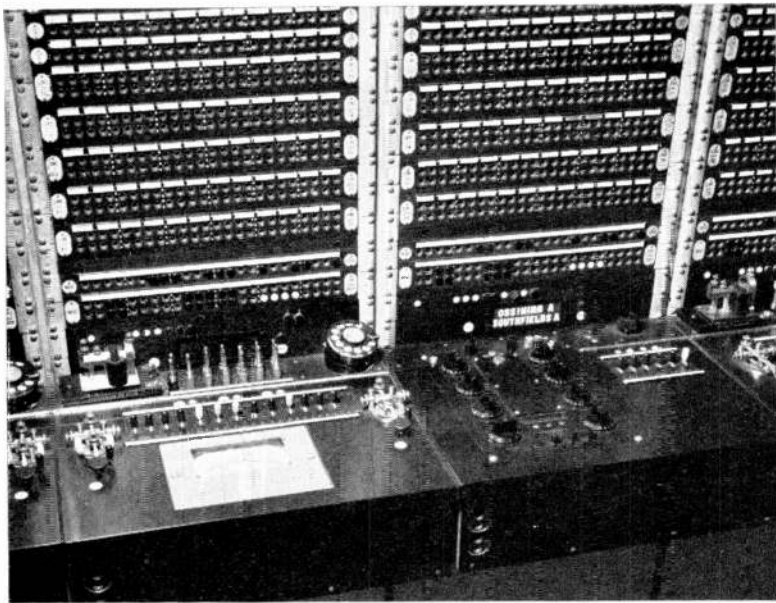
TESTSHELF OF NO. 5 TESTBOARD SHOWING VOLTMETER
AND WHEATSTONE BRIDGE TESTING ARRANGEMENT

Probably the most difficult resistance measurements are those of extremely low values. Examples of these are: the internal resistance of an ammeter (or the resistance of an ammeter shunt); the resistance of an electrical connection such as the connection between cells of a storage battery; the resistance of an electrical bond, such as bonds used to prevent electrolysis and connected between railroad rails and water pipes or from one railroad rail to another.

Where very low resistances are to be measured accurately, it is usually a complicated laboratory process. Fortunately, we have but few such cases in our work, though there are cases where the presence of low resistance values is to be determined but not necessarily with a great degree of accuracy. For example, in the case of a connection between the cells of a storage battery, we may desire to know whether the resistance of the connection is greater than it should be. Were this to be accurately measured, the measurement would be a difficult one to make but it can usually be determined for practical purposes by some simple test such as touching the two sides of the connection with the terminals of a telephone receiver and listening for a click due to the potential drop caused by the resistance. It follows that we may confine our attention here to the practical methods used for measuring either those resistance values which are appreciable, such as the ones that are important in simple circuits, or those resistance values which are extremely high, such as the insulation resistance of cable or open wire conductors.

Two basic methods for measuring electrical resistances are discussed below:

a short-circuit when connected across points in a circuit having a considerable difference in potential, while the voltmeter is for most practical purposes an open circuit, and unless connected to points having potentials higher than its greatest scale reading, it cannot be damaged from excess current values. In the language of the electrician, **the ammeter must always be inserted and never connected across.**

Voltmeters and ammeters are manufactured for different ranges of voltage and current values and one instrument often has several scales. Instruments for measuring small values are prefixed with milli, meaning one-thousandth, or micro, meaning one-millionth. Thus, we have milliammeter, millivoltmeter, etc. It is obvious that **an instrument must not be used when the value of the voltage or current to be measured is likely to be greater than the maximum scale reading.**

5.2 Simple Resistance Measurements

There are numerous methods for measuring electrical resistance and the one which is most practical depends upon—

(a) the magnitude of the resistance to be measured;

(b) the conditions under which it is to be measured;

(c) the degree of accuracy required.

FIG. 5-3 DROP IN POTENTIAL RESISTANCE MEASUREMENT

a. Voltmeter-Ammeter Method

Figure 5-1 shows a simple series circuit. Let us assume that it is desired to determine the value of the resistance R_1. We have learned that if a voltmeter is connected across the terminals a and b as shown, it will measure the potential drop across the resistance. But if, at the same instant this reading is taken, an ammeter is so inserted as to read the value of the current flowing through the resistance R_1, we will have not only an emf reading but a current reading as well and from the two, the value of the resistance may be calculated by Ohm's Law.

Example: In Figure 5-1, the voltmeter reading is 5 volts and the ammeter reading is .5 ampere; what is the value of resistance R_1?

Solution:

$$R_1 = \frac{V_1}{I_1} = \frac{5}{.5} = 10 \text{ ohms, ans.}$$

b. Drop in Potential Method

If in Figure 5-3, it is desired to determine the value of the resistor R_1, the "drop in potential method" can be used if a second resistor R_2 of known value is inserted in series and the voltage drops across both R_1 and R_2 are measured. Since the two resistors are in series, the same current is flowing through both and from Ohm's Law:

$$I = \frac{V_1}{R_1} \text{ , and also } I = \frac{V_2}{R_2}$$

Therefore,

$$\frac{V_1}{R_1} = \frac{V_2}{R_2}$$

which may be written, either—

$$\frac{V_1}{V_2} = \frac{R_1}{R_2}$$
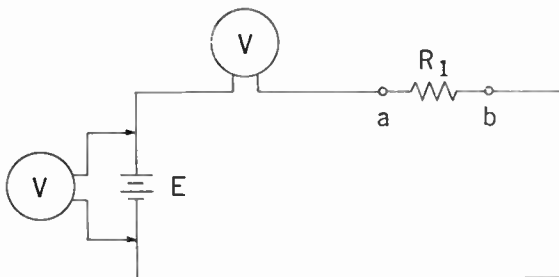
FIG. 5-4 INSULATION MEASUREMENT
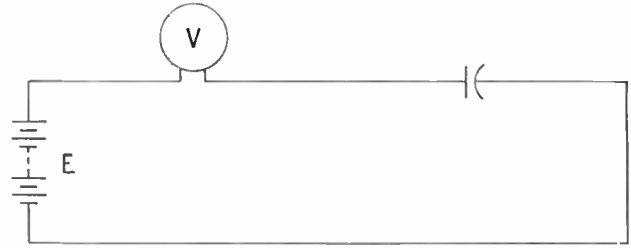
FIGURE 5-5

or

$$R_1 = R_2 \frac{V_1}{V_2} \qquad (5:1)$$

Example: If in Figure 5-3 the value of R_2 is 10 ohms and the drop across it is 12 volts, what is the value of R_1 which has a drop of 8 volts?

$$R_1 = R_2 \frac{V_1}{V_2} = 10 \times \frac{8}{12} = 6.67 \text{ ohms, ans.}$$

5.3 Insulation Measurements

The application of the drop in potential method which has greatest importance in telephone and telegraph work is its special adaptation to insulation measurements.

If the series circuit in Figure 5-3 contains no resistance other than R_1 and R_2, it is not necessary to measure the drop across R_1 because it will be equal to the potential of the battery minus the drop across R_2. The equation for this special case may then be written—

$$R_1 = R_2 \frac{E - V_2}{V_2} \qquad (5:2)$$

where E is the emf of the battery.

If R_1 is very high in value such as a "leak" due to poor insulation, it can be measured using equation (5:2) but instead of using a second known resistance, the voltmeter itself may be inserted in series with the battery and R_1 as shown in Figure 5-4. The reading V_2 then applies to the drop across the **voltmeter's own resistance** which, as has been previously stated, is very high. But since the resistance being measured is very high, this gives greater accuracy than if a known resistor R_2 having a lower value were inserted and a drop of lower value measured across it. As a matter of fact, voltmeters used for measuring insulation are especially designed to have abnormally high internal resistance; the ones used in the standard testboard testing circuits have a resistance of 100,000 ohms.

Figure 5-5 shows the drop of potential method with series voltmeter for measuring the insulation of a capacitor. Figure 5-6 shows a "leak" between two cable conductors and Figure 5-7 a "leak" between an open wire and ground, both being measured in the same manner.

For this application equation (5:2) is ordinarily written—

$$X = r\ \frac{E - V}{V}$$

or

$$X = r\left(\frac{E}{V} - 1\right) \tag{5:3}$$

where X is the unknown insulation resistance in ohms and corresponds to R_1, r is the resistance of
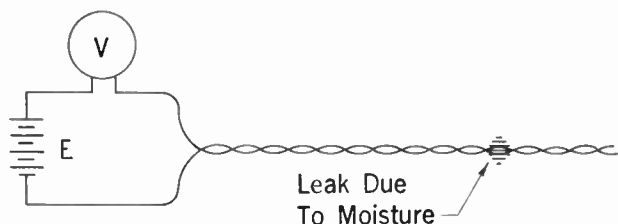
FIG. 5-6 METALLIC INSULATION TEST

the voltmeter and corresponds to R_2, E is the voltage of the battery and V is the voltmeter deflection.

Example: The voltmeter shown in Figure 5-5 has a resistance of 100,000 ohms. If it reads 8 volts as shown and 150 volts when connected directly across the battery terminals, what is the insulation resistance of the capacitor?

Solution:

$$X = r\left(\frac{E}{V} - 1\right)$$

$$= 100,000 \left(\frac{150}{8} - 1\right)$$

$$= 1,775,000 \text{ ohms, ans.}$$

Note: Insulation resistance is usually expressed in megohms instead of ohms on account of its normally high value. One megohm equals
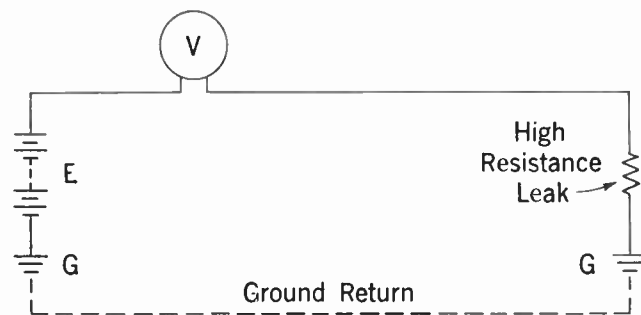
FIG. 5-7 TEST FOR INSULATION OF SINGLE WIRE

one million ohms. Equation (5:3) may accordingly be written:

$$X = r\left(\frac{E}{V} - 1\right) \div 1,000,000$$

where X is insulation resistance in megohms instead of ohms.

5.4 Theory of the Wheatstone Bridge

In practice the Wheatstone bridge provides a faster and more accurate means of making resistance measurements although the principles involved are not different from those of the potential drop method of measurement outlined in the preceding Article. Before reviewing specific Wheatstone bridge measurement techniques, how-
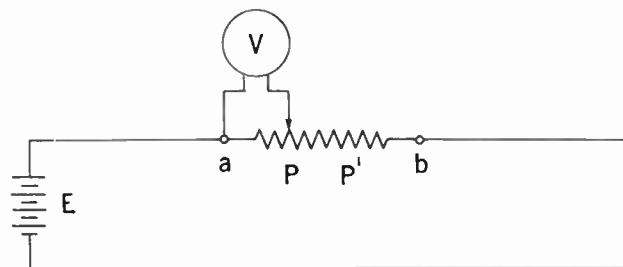
FIGURE 5-8

ever, it may be well to analyze briefly the broad general theory of this very widely used measuring instrument.

In Figure 5-8, the voltmeter has one terminal permanently connected to a and the other terminal may be moved along the resistor ab. The voltmeter reading will be zero when both terminals are at a, and will gradually increase as P is moved toward b. We shall find that the potential drop measured between the points a and P is

[32]
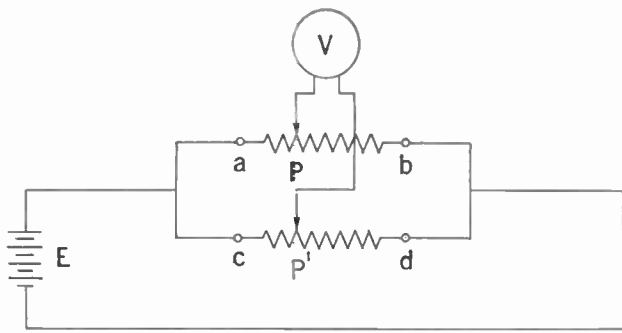
FIGURE 5-9

always proportional to that part of the resistance
between the two points, or we may write:

$$\frac{aP}{aP'} = \frac{V}{V'}$$

where V' is the potential drop measured between
a and any other point P'.

If instead of having one resistor as shown in
Figure 5-8, we have two parallel resistors as
shown in Figure 5-9, and one terminal of the volt-
meter is moved along resistor ab while the other
terminal is moved along resistor cd, we shall find
that when that part of the resistance ab between
the points a and P is proportional to that part of
the resistance cd between the points c and P', the
difference in potential between the points P and
P' will be zero and there will be no reading of the
voltmeter. Mathematically, this may be expressed:

$$\frac{aP}{cP'} = \frac{ab}{cd}$$

Likewise, we may develop a similar expression
for the remaining part of the resistance:

$$\frac{Pb}{P'd} = \frac{ab}{cd}$$

From these two relations, we may write:

$$\frac{aP}{cP'} = \frac{Pb}{P'd} \qquad (5:4)$$
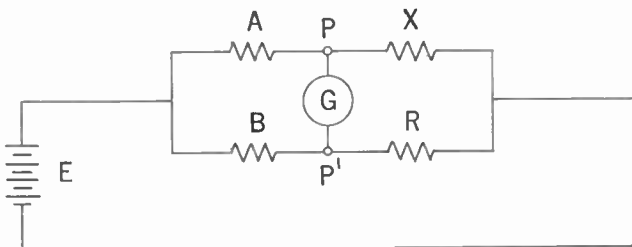
FIGURE 5-10

In Figure 5-10 let us assume that the resistances
represented by the branch A, the branch B, and
the branch R, are known, and that the resistance
shown as the branch X is unknown. Inasmuch as
the meter connected between the points P and P'
is merely being used to determine that these two
points have the same potential, a galvanometer
can be employed instead of a voltmeter, and will
be preferable in that it is more sensitive. If there
is no deflection in the galvanometer needle, we
may write the same relation as was given by
equation (5:4), namely:

$$\frac{A}{B} = \frac{X}{R} \text{ or } \frac{A}{X} = \frac{B}{R}$$

This equation can be expressed:

$$X = \frac{A}{B} R \qquad (5:5)$$

which is the usual equation of the Wheatstone
bridge.
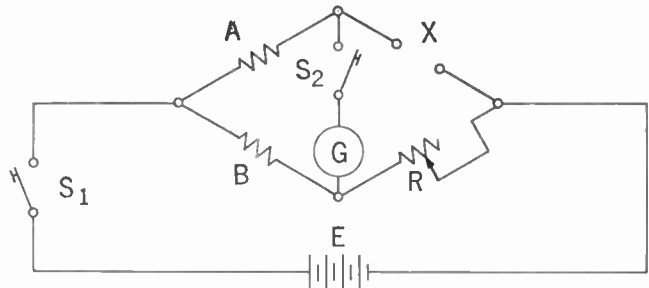
FIG. 5-11 STANDARD WHEATSTONE BRIDGE CONVENTION

Figure 5-11 illustrates the conventional method
of showing the Wheatstone bridge. It is almost
identical to the arrangement shown by Figure
5-10, but has the resistors connected in a diamond
shaped diagram. S_1 is a switch for disconnecting
the battery when not in use, and S_2 is a similar
switch for disconnecting the galvanometer. Bind-
ing posts are shown for connecting the unknown
resistance to be measured, which is usually desig-
nated as X. The resistors A and B are called the
ratio arms of the bridge and the resistor R is
variable so that for any unknown resistance X,
the value of R may be adjusted to obtain a per-
fect balance, or to bring the galvanometer needle
to the stationary or zero point on the scale.
Though Figure 5-11 shows the resistance branch
R as variable and the arms A and B as fixed, a
balance could also be obtained by changing the
ratio A/B in equation (5:5) instead of varying
the value of R.

[33]

the same, which means that opening or closing the distant end of the cable pair does not in any way affect the measurement, the cross is known to have zero resistance (dead crossed). If the measurement with the distant end of the cable pair crossed is lower in value than the measurement with the distant end of the cable pair open, the cross itself has some definite resistance value, and the location, instead of being .654 mile away, is some point between .654 mile away and the office. One way to determine the exact location in this case is to make loop measurements from each end of the cable pair, and to calculate an imaginary location from each measurement on the assumption of a zero cross. The location, when calculated from the measurement made at the office end, will be too far away, and when calculated from the measurement made at the distant end, will be too near the office. The actual loca-
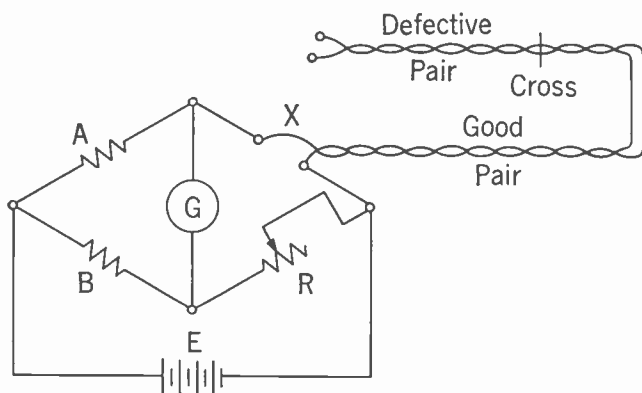
FIGURE 5-14

tion is the mean, or point half way between the two. Of course, in practice it may not be convenient to transfer the Wheatstone bridge to the distant end of the circuit in order to make measurements from that end. A substitute for this method, which amounts to the same thing, is to connect the distant end of the defective pair to a good cable pair as shown in Figure 5-14. This permits testing in both directions from the same office. If the exact length of the good pair is not known, it can be determined by making a measurement with the distant end crossed.

Note: As pointed out in the next Article, the quickest and most accurate method of locating a cross in practice is by the use of a Varley measurement. But the theory underlying the foregoing should be thoroughly mastered before taking up the later type of test.

Example: In Figure 5-14 the good cable pair when short-circuited at the distant end has a loop resistance of 63 ohms. When connected to the defective pair as shown, the measurement from the office to the cross over the good pair and the distant end of the defective pair is 108 ohms. The measurement from the office to the cross on the defective pair is 37 ohms. What is the distance in miles from the office to the defect, and what is the resistance of the cross if the cable pairs are 19 AWG?

Solution: Assume first that the cross has zero resistance. The imaginary distance from the office according to the measurement on the defective pair is as follows:

$$d_1 = \frac{37}{84.1} = .440 \text{ mile}$$

The length of the good pair is:

$$l = \frac{63}{84.1} = .749 \text{ mile}$$

The imaginary distance from the distant end is:

$$d_2 - l = \frac{108}{84.1} - .749 = .535 \text{ mile}$$

Then the actual cross is at point half way between .440 mile from the near office and .535 mile from the distant office or .749 − .535 = .214 mile from the near office. The actual location is therefore

$$\frac{.440 + .214}{2} = .327 \text{ mile, ans.}$$

The resistance of the cross caused an error in the single measurement location of .440 − .327 = .113 mile of cable pair. This expressed as resistance is .113 × 84.1 = 9.5 ohms, ans.

An important application of the simple loop resistance measurement is to determine any inequality in the resistance of individual conductors, or as is commonly expressed, to locate "resistance unbalances" due to such causes as defective splices in cable pairs or defective sleeve joints in open wire. This test, requiring at least three conductors, is ordinarily made by having the conductors crossed at the distant end and making measurements on various combinations:

Example: It is desired to determine the amount of resistance unbalance for the conductors of a phantom group between points A and B.

The Wheatstone bridge is located at A. The wires are designated 1, 2, 3 and 4 and all four wires are crossed at B.

Procedure: Measure the loop resistance with the bridge connected to wires 1 and 2. Let us assume a reading of 90 ohms.

Measure the loop resistance with the bridge connected to wires 2 and 3. Let us assume a reading of 90 ohms.

Measure the loop resistance with the bridge connected to wires 1 and 3. Let us assume a reading of 88 ohms.

Measure the loop resistance with the bridge connected to wires 1 and 4. Let us assume a reading of 88 ohms.

Solution:

$$W_1 + W_2 = 90 \qquad (a)$$

$$W_2 + W_3 = 90 \qquad (b)$$

$$W_1 + W_3 = 88 \qquad (c)$$

Subtracting equation (b) from equation (a) we have:

$$W_1 - W_3 = 0 \qquad (d)$$

Adding (c) and (d) we have

$$2W_1 = 88, \text{ or } W_1 = 44 \text{ ohms}$$

Substituting in (a) we have

$$44 + W_2 = 90 \text{ ohms or } W_2 = 46 \text{ ohms}$$

Substituting in (c) we have

$$44 + W_3 = 88 \text{ ohms or } W_3 = 44 \text{ ohms}$$

And since $W_1 + W_4 = 88$, we likewise have

$$44 + W_4 = 88 \text{ or } W_4 = 44 \text{ ohms}$$

Thus we learn W_2 has a resistance unbalance of 2 ohms, ans.

5.6 Varley Loop Tests

The Wheatstone bridge may be used to locate a defect due to a grounded conductor as well as a defect due to a cross between conductors. There are two recognized methods of making tests of this kind. One is known as the Varley loop test and is the more generally used; the other is known as the Murray loop test. Figure 5-15 shows the theory of an ordinary Varley loop test for a ground.

In comparing this figure with Figure 5-11, we can recognize a Wheatstone bridge circuit with the connections made in a little different way. The variable resistance R is in series with the resistance d of the defective wire from the office to the fault. The resistance X of Figure 5-11 becomes in Figure 5-15 the series resistance l of the good wire from the office to the distant end, plus the resistance x of the defective wire from the distant end to the fault. The battery connection is made through the ground to the fault itself. When a balance is obtained in this circuit, the value of R is equal to the loop resistance of the circuit from the defect to the distant end, if the A and B arms are equal. This may be seen by inspection, for it is evident that the adjustment of the R arm of the bridge is used merely to add resistance to the defective wire and since the resistance of the defective wire from the bridge to the fault balances an equal length of the good wire, the value of R when the bridge is balanced equals the resistance of the loop from the defect to the distant end.

Example: In Figure 5-15 the bridge is connected to a 30-mile circuit of 104 open wire. Each of the arms A and B is set at 1000 ohms. If the reading for the value of R is 22 ohms, how far is the ground from the office making the test?

Solution: Table III gives 10 ohms per mile for the loop resistance of 104 copper wire. The measurement of 22 ohms represents the resistance of the loop from the defect to the distant end. This distance is therefore $22/10 = 2.2$ miles. If the circuit is 30 miles long, the defect is located $30 - 2.2$ or 27.8 miles from the measuring office, ans.

The above example assumes that the two wires of Figure 5-15 are alike, and that the loop resistance values per mile given in the table are
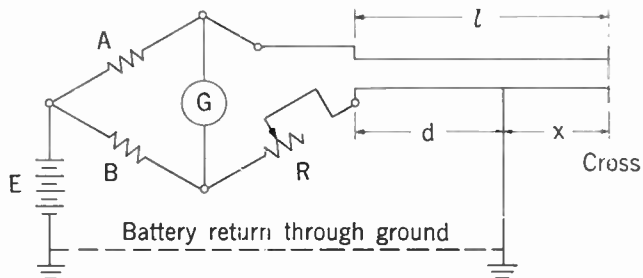
FIG. 5-15 GROUNDED VARLEY TEST

correct under all conditions. Although the first assumption will usually be true in practice, unit resistance values may vary appreciably due to temperature differences. In either event, it is still possible to locate the fault, by making an ordinary loop resistance measurement on the pair, in addition to the grounded Varley measurement.

Thus, referring again to Figure 5-15, it will be seen that when a Varley balance is obtained, with the bridge ratio arms equal—

$$R + d = l + x \qquad \text{(a)}$$

or

$$d = (l + x) - R \qquad \text{(b)}$$

This, of course, is true regardless of whether the good and the defective wires are of the same make-up.

Similarly, if the loop resistance L is measured, we have

$$L = l + x + d \qquad \text{(c)}$$

from which

$$d = L - (l + x) \qquad \text{(d)}$$

Now, adding (b) and (d), we get

$$2d = L - R \qquad \text{(e)}$$

In other words, **the loop resistance from the measuring end to the fault is equal to the loop resistance measurement minus the Varley measurement.** Since we do not know precisely the unit resistance value of the two wires, we still do not

know the exact distance to the fault. It is obvious, however, that if the wires are of uniform make-up throughout their whole length, the ratio of the distance to the fault to the total length of the line will be equal to the ratio of the loop resistance to the fault to the total loop resistance. That is, if we designate the distance to the fault as k, and the total length of the line as D, then—

$$\frac{k}{D} = \frac{2d}{L} \qquad \text{(f)}$$

from which

$$k = \frac{2d}{L} \times D \qquad \text{(g)}$$

or applying (e) above

$$k = \frac{L - R}{L} \times D \qquad \text{(h)}$$

Example: Assume as in the above example that the total circuit length is 30 miles and that the Varley reading is 22 ohms with the ratio arms equal. If a loop resistance measurement gives 300 ohms, what is the distance of the ground from the measuring end?

Solution:

$$k = \frac{300 - 22}{300} \times 30 = 27.8 \text{ miles, ans.}$$

A modification of the Varley test may be used for accurately measuring resistance unbalances, which is in some respects preferable to the method of combination loop measurements described in the foregoing Article. It is called the metallic Varley, and is shown by Figure 5-16(A). In making this test, all wires are short-circuited at the distant end in the same manner as when making a series of loop tests for the various combinations of wires. At the testing office, one wire of the combination is used for the battery return, instead of a circuit formed by grounding at the distant office. Two of the remaining wires are then connected to the bridge and R is adjusted to give a balance. If a balance cannot at first be secured, this indicates that the higher resistance wire is in series with R, and the connections to the bridge terminals are reversed. If the arms A and B are equal, the value of R then obtained represents the difference between the resistance of the two wires, and no calculations are required. When all combinations of wires are tested by the metallic Varley excepting the battery return wire, this wire
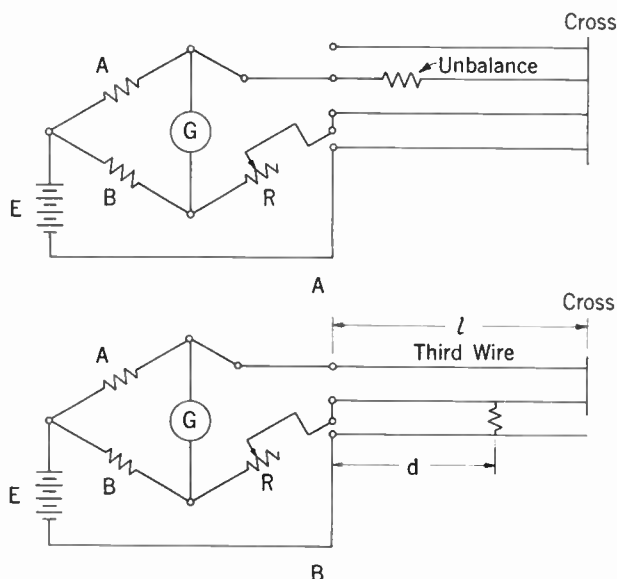
FIG. 5-16 METALLIC VARLEY TESTS

[38]

may be interchanged with any one of the others and included in the tests.

A similar test requiring only three wires is commonly used in testboard work for locating crosses, particularly those having high resistance. As noted in the preceding Article, the location of a cross having resistance by the use of loop resistance measurements involves certain difficulties. By using a good third wire of the same gage as that of the pair in trouble, and connecting the bridge for a metallic Varley measurement as shown in Figure 5-16(B), the resistance of the cross is removed from the "balanced" circuit of the bridge and placed in the battery circuit. Here it has no effect on the measurement, providing its resistance is not so high that the current supplied to the bridge is insufficient for its satisfactory operation. As may be seen from the diagram of connections, when the bridge is balanced with equal values in the ratio arms A and B, the resistance of the good third wire, plus the resistance of one wire of the crossed pair from the distant end to the fault, is equal to the resistance of one wire from the fault to the measuring end plus the resistance, R, in the rheostat arm of the bridge; or, we may write—

$$l + (l - d) = d + R$$

from which

$$d = \frac{2l - R}{2}$$

In locating a cross by this method in practice, it is only necessary to make a Varley measurement as described above and a loop resistance measurement on the pair consisting of the good third wire and one wire of the crossed pair, shorted together at the distant end. Then the loop resistance of the crossed pair from the measuring end to the fault may be obtained directly by subtracting the Varley reading from the loop resistance reading.

The Varley test may also be used for locating a cross between one wire of a circuit and some other wire of different characteristics, such as one wire of an iron circuit. The procedure here is to ground the wire of the second circuit, cross the first circuit at the distant end, connect the bridge to it and locate the ground by the Varley method described above, which is equivalent to locating the cross.

In practical line testing work, it is sometimes desirable to use all three of the types of measurements we have been discussing for locating a cross
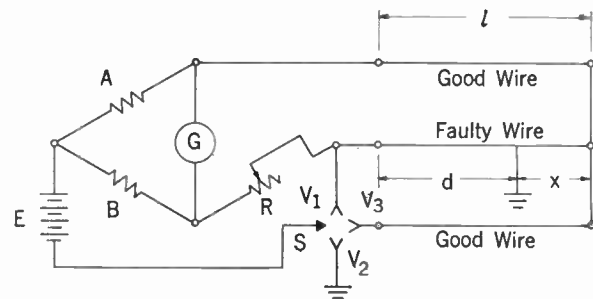
FIG. 5-17 3-VARLEY TEST FOR GROUND

or ground—i.e.: a grounded Varley measurement, a metallic Varley measurement, and a loop resistance measurement. This method, commonly called **the 3-Varley method** will give accurate results regardless of whether the several line wires used (including the defective wire) have the same overall resistance, regardless of the resistance of the cross or ground connection, and regardless of the insulaton resistance between the wires and between the wires and ground.

Figure 5-17 illustrates the arrangement for making such measurements. In this case, where a grounded fault is to be located, two good wires, which may or may not have the same resistance, and the faulty wire are connected together at the distant office. A metallic Varley measurement which will indicate the difference in resistance between one good wire and the faulty wire, a grounded Varley measurement which gives the resistance from the distant office to the fault, and a loop measurement giving the total resistance between the two offices are made.

The loop measurement is taken by connecting the switch arm s to the V_1 position (Figure 5-17) and adjusting R until the bridge is balanced. This reading will be designated as V_1.

The grounded Varley measurement is made by switching to the V_2 position and adjusting for balance. This reading is designated as V_2.

A metallic Varley measurement is made by switching to the V_3 position. This reading is V_3.

If the A and B arms are set at equal values for all three measurements, we can see by inspection of Figure 5-17 that

$$V_1 = l + x + d$$
$$V_2 = l + x - d$$
$$V_3 = l - x - d$$

Now we may note that

$$V_1 - V_3 = 2d + 2x$$
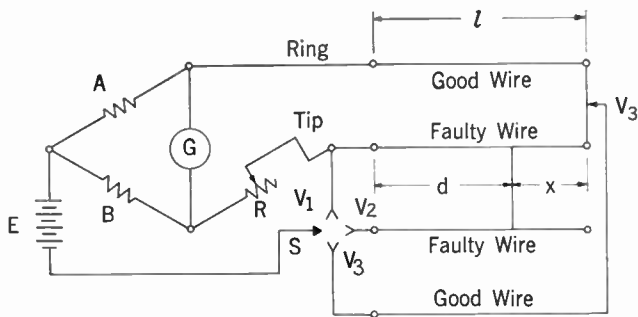
FIG. 5-18 3-VARLEY TEST FOR CROSS

which is twice the resistance of the faulty wire. Similarly

$$V_1 - V_2 = 2d$$

which is twice the resistance from the home office to the fault location. Then calling the distance to the fault k and the total length of the wire D as in (h) preceding, it is evident that

$$k = \frac{V_1 - V_2}{V_1 - V_3} \times D$$

It is particularly important to note in connection with this 3-Varley measuring method that the measured values in all cases include only values of the resistance of the faulty wire itself. The specific resistance values of other wires used in making the measurements are immaterial since they do not appear in the final equations.

Example: Assume that in Figure 5-17, a toll circuit is 42 miles long and that the following 3-Varley measurements are obtained:

$$V_1 = 3970$$

$$V_2 = 1700$$

$$V_3 = \quad 20$$

What is the distance from the measuring office to the fault?

Solution:

$$k = \frac{V_1 - V_2}{V_1 - V_3} \times D$$

$$= \frac{3970 - 1700}{3970 - 20} \times 42 = 24.15 \text{ miles}$$

The 3-Varley method is equally applicable for locating faults due to crosses where the measuring set-up would be as shown in Figure 5-18 and the calculations for determining the location would be the same as in the case of the ground fault discussed above.

5.7 Murray Loop Tests

The theory of the Murray loop test is similar to that of the Varley. But instead of setting the arms A and B to have equal values and using the adjustable dials R to compensate for the difference in wire resistance between the good wire connection and the defective wire connection, the arm B is eliminated altogether and the variable resistance arm is connected in its place as shown in Figure 5-19. In this arrangement, the ratio of the reading R to the setting of the arm A is equal to the ratio of the resistance of the defective wire from the measuring office to the ground to the resistance of this same wire from the ground to the distant office plus the resistance of the good wire, or expressed mathematically

$$\frac{R}{A} = \frac{l - d}{l + d}$$

This, of course, assumes that the defective and good wires have the same series resistance per mile, as would ordinarily be the case where for any given circuit being tested the defective wire's mate is used.
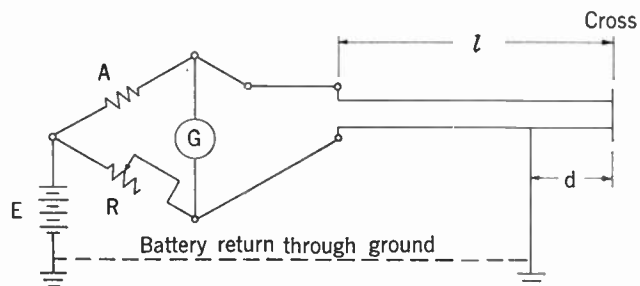
FIG. 5-19 MURRAY TEST

Example: In Figure 5-19, the arm A is set at 1000 ohms, and the bridge is balanced by varying the arm R. If the value of R is 634 ohms and the length of the circuit under test is 65 miles, what is the distance from the testing office to the fault?

Solution: The simple bridge relation gives

$$\frac{R}{A} = \frac{l - d}{l + d}$$

or

$$\frac{634}{1000} = \frac{(65 - d) \times \text{res. per mile}}{(65 + d) \times \text{res. per mile}}$$

If the resistance per mile of each wire is the same, this factor will cancel and we have—

$$\frac{634}{1000} = \frac{65 - d}{65 + d}$$

which gives by cross multiplying

$$634 \, (65 + d) = 1000 \, (65 - d),$$

or

$$1634d = 23790$$

from which

$$d = 14.56 \text{ miles}$$

$$l - d = 50.44 \text{ miles, ans.}$$

The advantage of the Murray test in locating a fault such as discussed above lies in the fact that the test does not require the use of a third wire as would be necessary in the Varley method. Except in certain special conditions involving rural lines, one pair service cables, etc. the Murray test is rarely used in telephone practice for locating grounds or crosses. The Murray type connection is commonly used, however, for locating opens. But since the wires here are open, it is obvious that no ordinary d-c measurement can be made. Instead, a low frequency a-c is generated by means of an "interrupter" which reverses the battery voltage 8 times per second and simultaneously reverses the polarity of the galvanometer connections. The bridge when balanced then compares the "capacitance" of the good wire to its far end with that of the defective wire to the point where it is open.

As we shall see in later Chapters, the capacitance value of the wire is directly proportional to its length and the ratio given by the bridge reading will therefore be equal to the ratio of the length of the good wire to the length of the defective wire to the fault.

There are a number of other standard tests made with the Wheatstone bridge and with these as well as with the tests that have been described, the procedure in practice is somewhat more involved than the simple theory might indicate. There are in nearly all practical tests various complicating factors such as temperature variations, effect of loading coils, short lengths of cable, irregular facilities, etc., which must all be considered if accurate locations are to be made.

For instance, in many toll cables, parts of each section are aerial and other parts are underground. There is normally a considerable temperature difference between aerial and underground facilities and a corresponding difference in the wires contained in the cable sheaths. In such wires, it is necessary to apply correction factors to the measurements taken in order to locate a fault accurately. The details of how these various factors are taken care of in practice present a rather complete study in themselves, however, and their consideration is beyond the scope of this book. The intent here has been only to treat a few of the outstanding testing methods in a more or less theoretical way, with a view to establishing the general principles upon which all testing work is based.

CHAPTER 6

THE DIRECT-CURRENT DYNAMO-ELECTRIC MACHINE

6.1 Induced Electromotive Force

Chapter 4 describes how lines of magnetic induction exist around any wire in which there is an electric current. Not only does a current establish such a field, but conversely a magnetic field can be made to create an electromotive force. Voltage may be induced in any conductor by moving it through a magnetic field in such a manner that it "cuts" the magnetic lines. If the wire indicated in cross-section by the circle in Figure 6-1 is moved horizontally to the right through the magnetic lines having a direction vertically downward, it may be considered that the wire displaces or "stretches" the lines, which may be thought of as possessing a certain elasticity. This finally causes them to wrap themselves around the conductor, as shown. Referring to Figure 4-11 in Chapter 4 and applying this figure conversely to our new conditions, we find that a magnetic field which loops around a conductor in a clockwise direction, gives rise to a current flowing "into" the conductor as seen in cross section. This is illustrated in Figure 6-1(D) and (E).
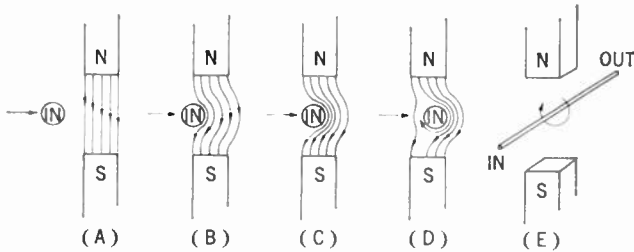
FIG. 6-1 WIRE MOVING THROUGH MAGNETIC FIELD

This rule, stated in another way, is called the **right-hand rule** for remembering the induced emf relation. It is illustrated in Figure 6-2. The forefinger of the right hand represents the direction of the lines of magnetic induction (flux—north to south) ; the thumb, when pointed perpendicular to the forefinger, represents the direction in which the conductor moves; and the second finger, when perpendicular to both the forefinger and the thumb, gives the direction of the induced emf, or the direction of current flow. If a galvanometer

is connected to the conductor, as in Figure 6-3, it will be found that the effect is more noticeable when the conductor is moved swiftly. From these
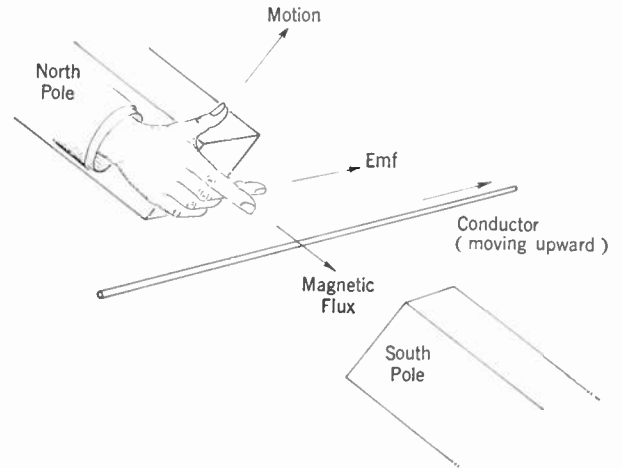
FIG. 6-2 RIGHT-HAND RULE

and other similar experiments we learn that the law for induced emf may be stated as follows :

When any conductor is made to cut lines of magnetic induction there will be an emf induced in it, and the direction of the emf, the direction of the flux, and the direction of the motion of the conductor have a perpendicular relation as shown by the right-hand rule. The magnitude of the in-
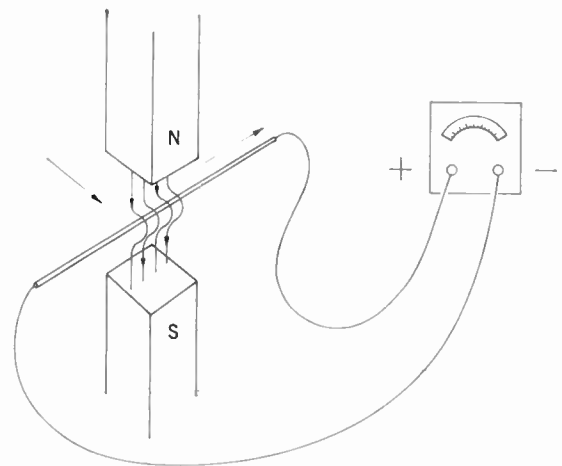
FIGURE 6-3
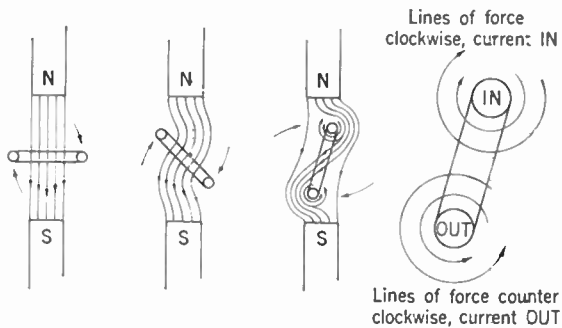
Fig. 6-4 Loop Rotating in Magnetic Field

duced emf depends upon the rate of cutting mag-netic lines, or the number of lines cut per second.

6.2 Emf Induced in a Revolving Loop

Instead of a single conductor cutting lines of magnetic induction, we may have a loop of wire revolving in the magnetic field between the poles of a magnet, as shown in Figure 6-4. In this case, the conductor nearest the south pole moves to the left while the conductor nearest the north pole moves to the right, and the emf induced has a different direction in the two conductors.

But, because the loop is complete, these emf's will aid in causing a continuous current in the direction a-b-c-d, as shown by Figure 6-5 (A). The values of these emf's at any given instant will de-pend upon the position of the loop. When the plane of the loop becomes perpendicular to the magnetic field as in Figure 6-5 (B), each conduc-tor will be moving parallel to the direction of the lines, the loop will be in a neutral position, and the generated emf will have decreased to zero. If the loop is then turned through an angle of 90° in the same direction (Figure 6-5 (C)), it will again be cutting lines at the maximum rate, but

(A)   (B)   (C)
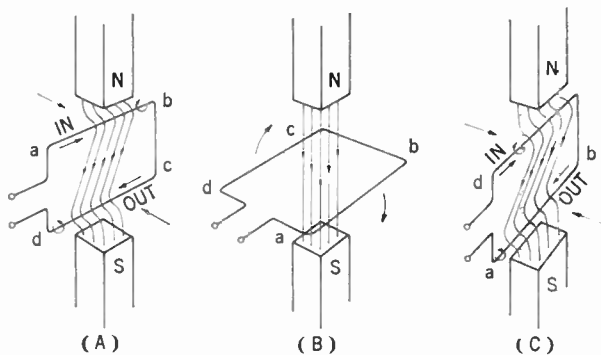
Figure 6-5

the emf will be reversed with respect to the loop itself and the current will be in the direction d-c-b-a, or opposite to that in Figure 6-5 (A).

With the loop revolving at constant speed, the emf induced in it and the resultant current are proportional to the number of lines cut, which in turn is proportional to the horizontal motion of each conductor of the loop. The maximum emf is induced when the plane of the loop is parallel with the lines, and the minimum (zero) when it is perpendicular to the lines. At every intermediate point, the value of the emf may be determined by the horizontal motion of the loop per angular degree through which it turns.

Figure 6-6 shows a mechanism which illustrates the way in which this current varies. The wheel d is rotating at a constant speed, causing the at-tached pin a to slide in the slot c, moving the bar b (with the pencil e attached) vertically between the guides gg. When the horizontal component of the motion of the pin a is a maximum, that is,
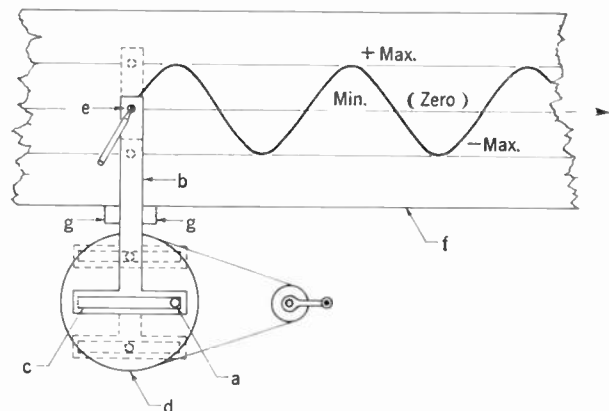
Fig. 6-6 Mechanism for Drawing Sine Wave

when the motion is in an entirely horizontal direc-tion, the pencil e is at either its highest or lowest position, depending upon whether the motion of a is from left to right or right to left. When the hor-izontal motion of a is zero, e is midway between its extreme high and low points. If f repre-sents a strip of paper which is being moved hori-zontally to the right at a constant speed, the pen-cil e will draw a curve as shown. This curve will indicate a positive maximum (or highest point) when the horizontal motion of a to the right is a maximum; and will indicate center or zero points when the horizontal motion of a is zero. If the pin in this mechanism represents one conductor of a loop of wire revolving in a vertical magnetic field,

| 43 |
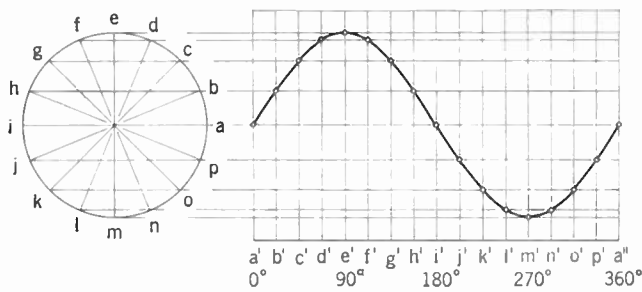
FIG. 6-7 GRAPHICAL CONSTRUCTION OF SINE WAVE

the position of the pencil *e* with respect to the mid-point of its travel represents the emf induced in the conductor. This is apparent since the induced emf in each loop is proportional to the horizontal motion of the loop. The curve not only represents maximum and zero points but shows all intermediate values of the induced emf as well.

Such a curve is called a **sine wave**. It is the **fundamental** wave form in alternating-current circuits of all kinds. A sine wave may be actually plotted by the method shown in Figure 6-7, where the horizontal lines are continuations of points *a*, *b*, *c*, etc., and the vertical lines *a'*, *b'*, *c'*, etc. are equally spaced and indicate **angular degrees of rotation**. The intersections of lines *a* and *a'*, *b* and *b'*, etc. indicate points on the sine curve.

6.3 Principle of the Direct-Current Generator

The revolving loop or **armature** shown in Figure 6-5 may be connected to slip-rings, as shown in Figure 6-8 (A). In this case the resulting emf be-
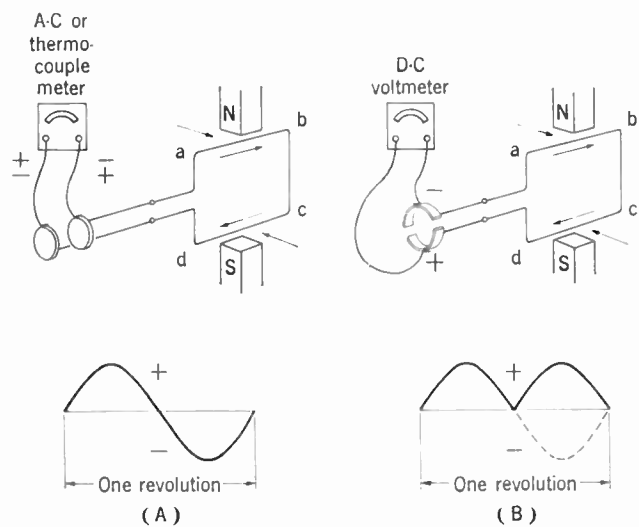
FIG. 6-8 SLIP-RINGS AND COMMUTATOR

tween the two terminals or **brushes** will reverse in direction as the loop revolves, giving rise to an **alternating** emf, one cycle of which is plotted in the Figure. If it is desired to produce a unidirectional emf, it is necessary to devise some means for reversing the connections to the loop at the same time that the current in the loop reverses. This is done by means of the **commutator** shown in Figure 6-8 (B). This commutator reverses the connections to the armature leads just as the emf or current is reversed, changing the negative half-cycle to a positive pulsation. The resultant emf then consists of two positive pulsations per revolution of the loop, as shown.

Generators may be constructed with more than one loop, as in Figure 6-9 in which two loops and four commutator segments are shown. The resultant emf is represented by the full lines at the right of the Figure. Comparing Figure 6-8 with Figure 6-9, it may be seen that an increase in the number of loops causes a smaller fluctuation in the armature emf. An armature wound with many turns therefore produces a practically con-
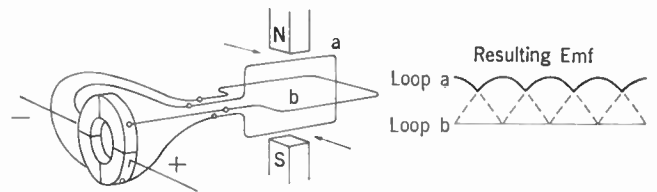
FIG. 6-9 EFFECT OF ADDITIONAL LOOPS

tinuous non-pulsating emf, causing a **direct** current to flow when the external circuit is closed.

In Figures 6-5, 6-8 and 6-9, we have assumed that the generator is equipped with permanent magnets which create the magnetic field. This is the case for small magnetos, but for other generators this field is furnished by electromagnets which are energized by a **field winding**. Direct-current machines are classified by the different means adopted to energize or "excite" this field winding. A **separately excited generator**, with the standard convention for indicating it, is illustrated in Figure 6-10. It is so called because the direct current through the field winding is furnished by an external source, such as another generator or a storage battery.

A more usual type is the **self-excited generator** which may be **shunt wound, series wound** or **compound wound.** The different methods of construction are shown schematically in Figure 6-11. As

[44]
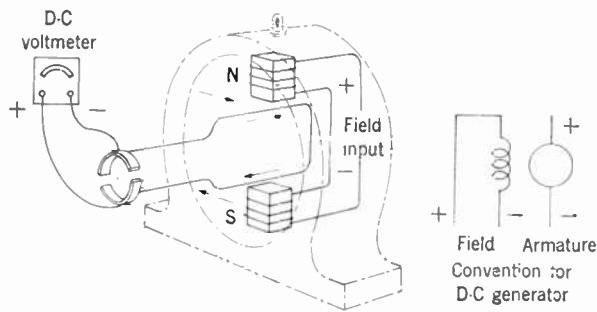
FIG. 6-10 SEPARATELY EXCITED D-C GENERATOR

the emf induced in the armature is proportional to the magnetic flux, which in turn is proportional to the current in the field windings, a variation in the field current will cause a change in the armature emf. With the shunt wound generator, an increase in load current causes a decrease in field current, as may be seen from a study of parallel resistances, and consequently causes a decrease in armature emf. On the other hand, in the case of the series wound generator, the armature emf increases with the load current. The compound wound generator is designed to neutralize this change in armature emf by balancing the series effect against the shunt effect. An **over-compounded generator** is constructed with the series effect predominant so that the voltage increases slightly with change from no load to full load. Figure 6-12 shows curves representing armature voltage plotted against load for these various types of generators.

Generators may be further classified by the number of poles, a four-pole machine being represented by Figure 6-13. In every case there are the same number of brushes as poles, alternate brushes being connected together, as shown, to form the armature terminals. The voltage with

FIG. 6-11 CONVENTIONS FOR STANDARD TYPES OF D-C GENERATORS

four poles will be double that with two poles if the same armature winding and the same machine speed are used.

6.4 D-C Generators for Supplying Central Office Power

Direct-current generators are widely used in telephone and telegraph work for supplying the several voltages required to operate the central offices. These include 24 and 48 volts used for "talking battery" and for operating certain relays and other apparatus; also various higher voltages, ranging up to a maximum of 130 volts, for electron tube plate supply and operation of telegraph circuits, etc. The motors which drive the generators are ordinarily supplied with power from com-

FIG. 6-12 LOAD-VOLTAGE CHARACTERISTICS OF STANDARD TYPES OF D-C GENERATORS

mercial power lines and to guard against the possible failure of this supply, storage batteries are always provided in central offices. These batteries are kept charged by the central office generators so that they can take over the load temporarily in case of failure of the primary power supply. Being always connected to the load, the storage batteries also have an important filtering effect in reducing noise caused by the generators.

There are several arrangements of generators and batteries that have been used in the past to develop central office power. The practice at present is to supply the load current continuously from one or more generators operated in parallel with each other and with a single storage battery. In this arrangement, the storage battery is "floated", or connected continuously across the main bus-bars. The normal generator voltage is then maintained at a value sufficiently high to

FIG. 6-13 FOUR-POLE GENERATOR

take care of the load requirements and to supply a small "trickle" charge to the battery, thus keeping it fully charged.

In power plants where a battery, continuously floated across the line, is used, it is desirable for the generator to have a load-voltage characteristic as nearly flat as possible. Because of its drooping characteristic, the shunt wound generator is therefore not suitable unless its voltage is constantly controlled by manual or automatic means.

The required flat characteristic could be obtained from an ordinary compound wound generator of proper design; but such a machine is not safe to use because in case of failure of the outside power, the floating battery would run the generator as a motor if the reverse current circuit-breaker failed to operate. In this situation, the reversed current in the series winding would cause the generator to operate like a series motor and because it would be carrying no load, it would tend to run at a dangerously high speed.

The present practice is to use regulated voltage rectifiers for continuous float power plants where the load requirements are small, and shunt wound generators with automatic voltage regulation for power plants of larger capacity.

6.5 Typical Central Office Power Plant

Figure 6-14 is a schematic of a typical power plant where the total load requirements are in the order of 100 to 4,000 amperes and both a generator and a rectifier are used as sources of supply. The battery is continuously floated in this type of plant and the charging unit voltage must, therefore, be maintained at a constant value. As indicated in the Figure, this is accomplished automatically by means of a motor-driven field rheostat associated with the shunt wound generator,
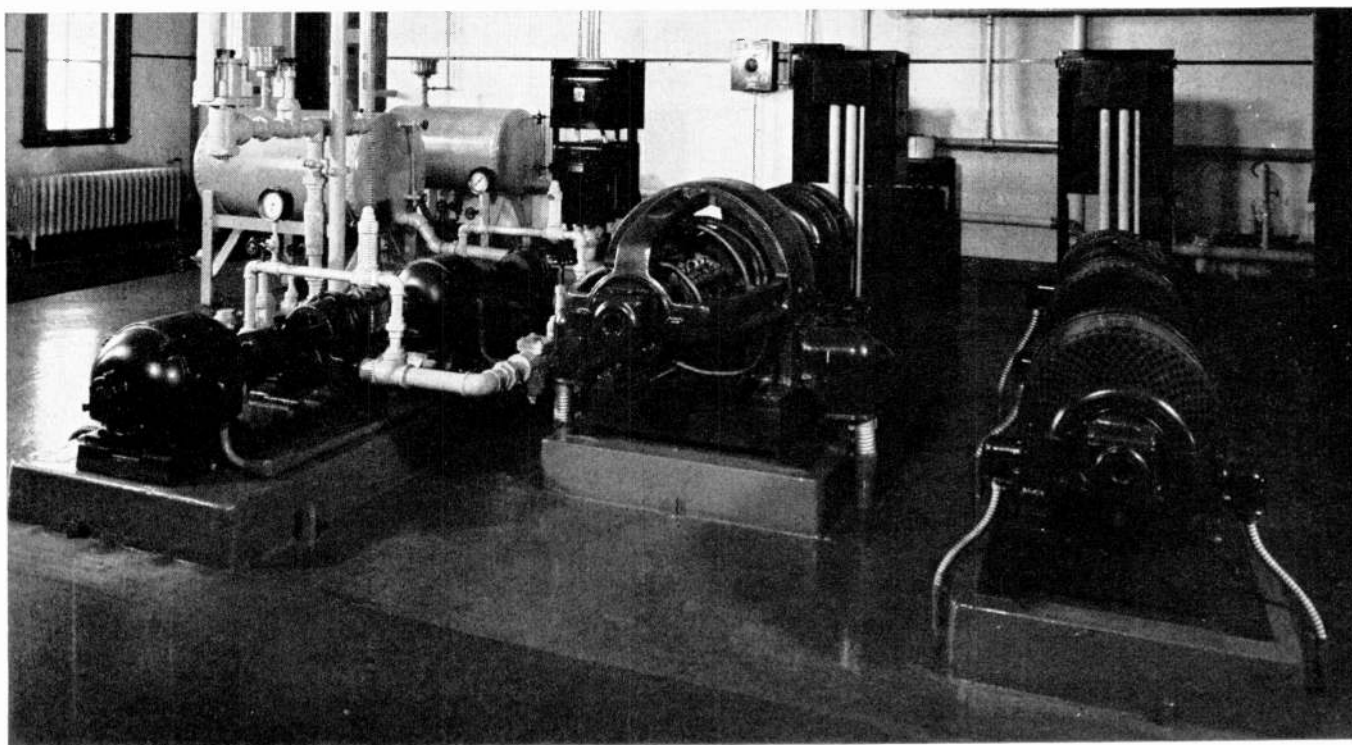
POWER PLANT AT OVERSEAS TRANSMITTING STATION
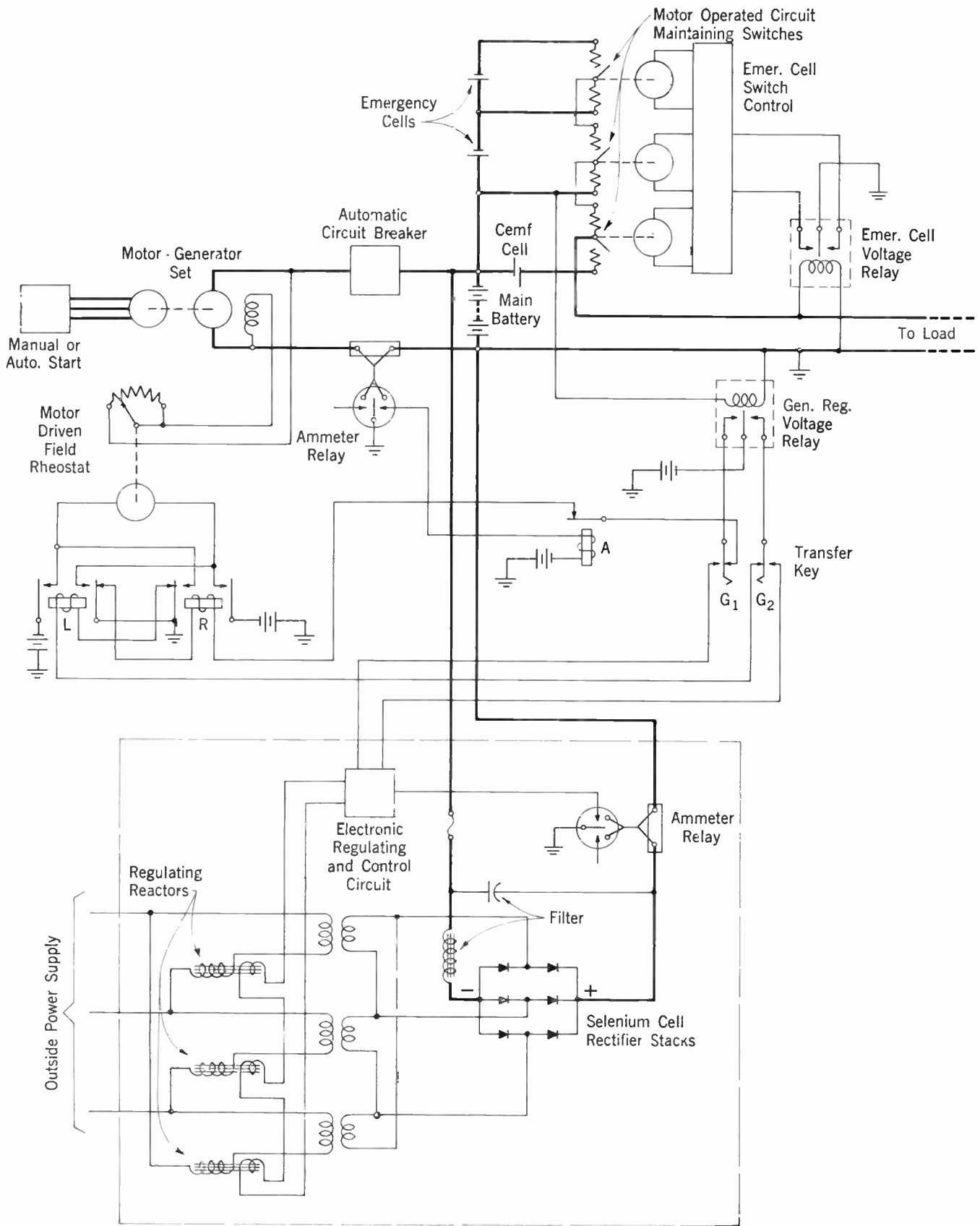
FIG. 6-14 TELEPHONE OFFICE POWER PLANT

[47]
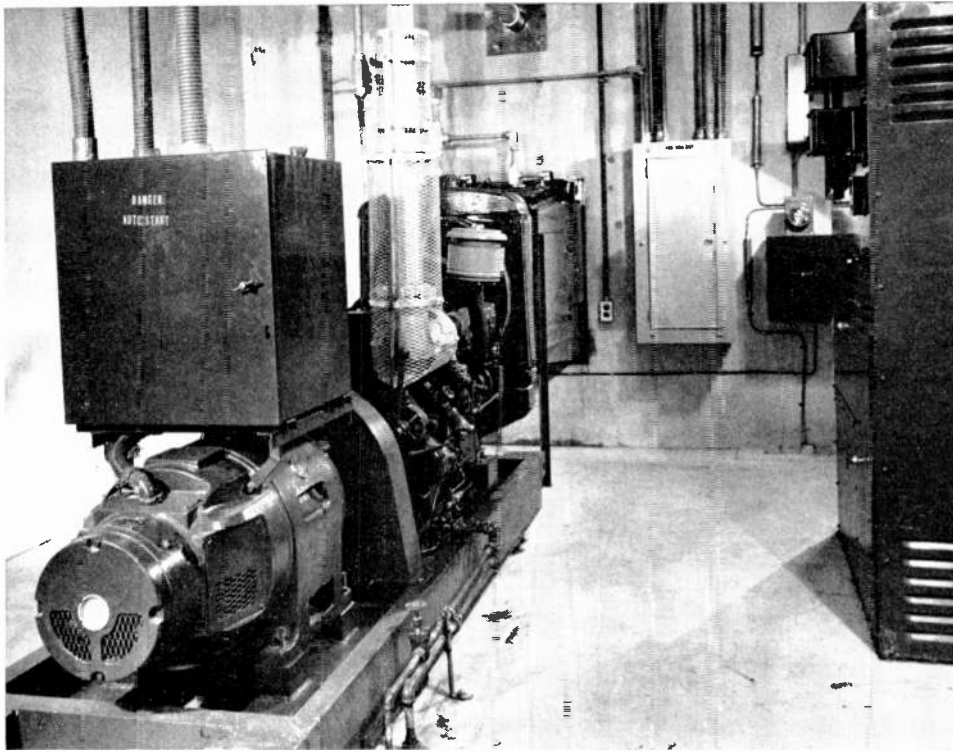
AUXILIARY POWER PLANT AT RADIO RELAY STATION

fore, inoperable. In such a case the load must be carried by the batteries alone and if the failure persists for an appreciable time the battery voltage will decrease below the required value. The emergency cells are then automatically cut into the circuit by means of a voltage relay bridged across the line, as shown.

Switches are provided for charging the emergency cells from the charging units in series with the main battery. However, since the emergency cells are not normally in use they are continuously supplied with a small trickle charge furnished by a copper-oxide rectifier (see Article 7.5) which normally maintains them in a fully charged condition. These switches and the copper-oxide rectifier are not shown in the drawing.

and by an electronic regulating and control circuit associated with the rectifier. A voltage relay (designated Gen. Reg. Voltage Relay in the drawing) is bridged directly across the main battery. As long as the battery voltage remains at its proper value, this relay is not operated; but if the battery voltage becomes too high, or too low, one or the other of the two relay contacts is closed. This causes either relay L or relay R to operate and the operation of either of these relays causes the motor-driven field rheostat to move in the direction which will restore the generator voltage to its normal value; or, in the case of the rectifier, causes the electronic control circuit to raise or lower the rectifier voltage as required. To avoid the possibility of overloading the charging unit, an ammeter relay is inserted in series with the line. When the unit is fully loaded, a contact on this relay closes causing relay A to operate and open the regulating voltage relay circuit. This prevents any further attempt on the part of the relay to increase the charging output.

The circuit includes two emergency cells which are connected to switches in such a way that one or both may be connected in series with the main battery. These cells are provided to take care of emergency conditions where the outside power supply fails and the charging units are, there-

The main battery is, of course, kept in a continuously charged condition as long as the plant is operating normally. When failure of the outside supply requires the battery to carry the load for an appreciable time, however, the battery will become more or less discharged and will therefore require special charging. In order to provide charging current in such a case, it is necessary to increase the output voltage of the charging unit above its normal value. But since the charging unit is connected directly to the load, an increase in its output voltage would also increase the load voltage. To avoid this, the circuit includes a **counter-emf** cell which is automatically inserted in series with the load circuit when the output voltage of the charging unit is increased above its normal value.

The cemf cell has the property, when current flows through it, of setting up a voltage opposing the voltage which is driving the current. The counter voltage is approximately 2 volts per cell and is substantially constant under wide variations of current. Physically, the cemf cell consists of two plates of pure nickel immersed in a caustic soda solution. As in the case of storage

batteries, which are discussed in Article 7.3 following, the size of the nickel plates depends upon the amount of current which the cell is required to handle. The cells are usually mounted along with the storage battery cells.

Figure 6-14 shows one generator and one rectifier but additional charging units may be included as required. To insure continuity of operation, a practical power plant always includes at least two charging units, and as many more may be added as are necessary to handle the maximum load. When the first unit becomes fully loaded the second is put into operation by throwing the transfer key shown on the drawing. Additional charging units when required are connected across the main leads to the battery but are manually controlled.

This power plant may be arranged so that the charging units will start automatically upon restoration of the outside power supply after failure. By including additional relay circuits, this general type of plant may also be arranged so that the needed units will be automatically started and connected to the line as the load increases and automatically disconnected and stopped as the load decreases.

6.6 Direct-Current Motors

When a conductor carrying an electric current is placed in a magnetic field at right angles to the lines of magnetic induction, there is a reaction between the circular field about the conductor and the field in which it has been placed. This reaction causes the lines set up by the two fields to aid or increase in number on one side of the conductor and to oppose or decrease in number on the other side. This gives the conductor a tendency to move across the magnetic field in a direction which depends on the direction of current flow in it.

If the conductor is a loop and is free to rotate, as in Figures 6-8 (B), 6-9 and 6-10, illustrating

d-c generators, it will revolve as a motor. In fact any d-c generator may be used as a motor if the current flows into the armature and field instead of out of the armature.

The direction of rotation may be determined by the left-hand rule where the left thumb represents direction of motion, the fore-finger direction of flux, and the middle finger direction of current flow.

When a motor is running, the armature conductors cut lines of magnetic induction, and an emf with a direction opposite to that of the applied emf is induced. This is called the counter-electromotive force, and the current in the armature is—

$$I = \frac{E_i - E_c}{R} \qquad (6:1)$$

where E_i is the impressed emf, E_c is the counter emf, and R is the armature resistance.

Since there is low cemf until the motor has reached about its normal speed, it will draw a very large current at starting unless this is prevented by a starting rheostat. This is a variable resistance placed in series with the motor's armature which is gradually cut out as the motor is brought up to its normal speed. A starting rheostat of some type must be used for all large motors, but is sometimes not required for small machines on account of the comparatively high resistance of their armatures.

The following are a few simple rules which have practical application to the use of motors:

1. The direction of rotation of a d-c motor may be reversed by reversing either the armature or field connections but not by reversing the supply leads.

2. The speed of a shunt wound motor may be adjusted by varying the field current. A decrease in field current gives an increase in speed, and vice versa.

3. A series motor must either have an increasing load with increase in speed, such as a fan, or its operation guarded by an attended controller; otherwise it will "run away".

7.3 Storage Batteries

A chemical battery that is capable of storing electrical energy delivered to it from some other source, and delivering this energy to an electric circuit at some later time, is called a storage battery. There are three principal types of storage batteries—the lead-acid, the Edison (iron-potassium hydroxide-nickel) and the nickel-cadmium. Because of its low internal resistance and more constant terminal voltage, the lead-acid type most nearly meets the exacting requirements for general telephone central office needs. (These requirements are discussed in Article 7.4 following.)

When the lead-acid cell is in a fully charged condition, the active constituents are a positive plate of lead peroxide (PbO_2) and a negative plate of spongy lead (Pb) in a dilute solution of sulphuric acid ($H_2SO_4 + H_2O$). When the battery is discharging, the current, passing from the positive to the negative plate through the external circuit, must return from the negative to the positive plate through the dilute acid (electrolyte). In doing so, it breaks the electrolyte into its component parts resulting in first, the spongy lead of the negative plate combining with the positively charged component (SO_4) of the electrolyte, forming lead sulphate ($PbSO_4$) and losing its negative charge; second, the oxygen of the lead peroxide of the positive plate combining with a part of the hydrogen liberated from the electrolyte, forming water, and converting the positive plate to pure lead; and third, a similar breaking up of the sulphuric acid at the positive plate, forming more water and converting some of the lead of the positive plate into lead sulphate by the same chemical action that takes place at the negative plate.

When the storage battery is charging, this chemical action is reversed. The charging current, in passing through the electrolyte in the opposite direction to that of the discharge current, breaks down some of the water of the electrolyte into hydrogen and oxygen. The oxygen travels against the current to the positive plate where it combines with the lead sulphate of that plate to form lead peroxide. The sulphate (SO_4) released by this action combines with hydrogen to form sulphuric acid. At the same time, hydrogen, traveling with the current to the negative plate, combines with the lead sulphate of that plate to form sulphuric acid. This leaves pure metallic or sponge lead on the negative plate, and the two plates and the electrolyte are thus gradually restored to their original charged condition.

The following chemical equation may be used to explain the action of discharge when reading from left to right and the action of charge when reading from right to left:

$$PbO_2 + Pb + 2H_2SO_4 \rightleftarrows 2PbSO_4 + 2H_2O \qquad (7:1)$$
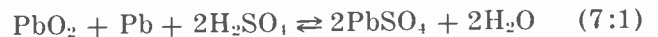
Storage batteries are built in a wide variety of sizes to meet the various load requirements. The capacity of a cell naturally depends on the total area of plate surface which is exposed to the electrolyte. The smallest cell consists of a single pair of plates having a total area of only a few square inches, while the largest cells may have more than 100 plates, each with an area of more than three square feet.

In modern central office practice, the plates of the smaller storage cells (up to a maximum ampere-hour capacity at an 8-hour discharge rate in the order of 1000) are mounted in sealed rubber containers. The containers are provided with vents to permit the release of gas and the plates are connected to terminals which project through
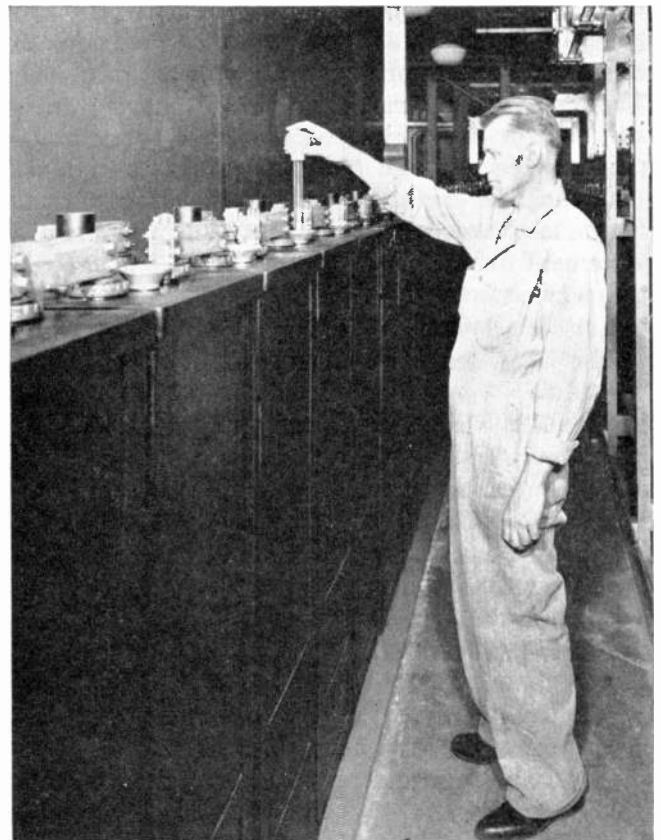
Enclosed Tank Type Storage Battery

[52]

the top. Larger cells are factory assembled in hard rubber tanks having sealed rubber covers as shown in the accompanying photograph.

In the practical operation of the lead-acid battery, we must be able to determine the state of charge or discharge at any time. It is not convenient to do this by chemical analysis, but in the foregoing explanation of the cycle of charge and discharge, there are two changes taking place that may be easily determined. One is the change in the electric charge held by the plates, resulting in a change in the emf of each cell. The other is the increase on discharge, and the decrease on charge, of the amount of water contained in the electrolyte, which increase or decrease, as the case may be, changes the specific gravity of the electrolyte. This latter condition gives the better index to the cell's operation and is the one ordinarily used.

The Edison storage battery has little application in telephone work. The more recent nickel-cadmium type of battery, however, is expected to become increasingly useful for such particular purposes as engine starting. It has discharge characteristics similar to the lead-acid battery but is not affected by short circuits and may be left in any state of discharge without detriment to the plates because the electrolyte (potassium hydroxide and distilled water) does not enter into the plate reactions but serves virtually only as a conducting medium. At high rates of discharge, as in engine starting, the voltage drops off less than for an equal capacity lead-acid battery. In consequence nickel-cadmium batteries for this service can be of somewhat less capacity.

7.4 Power Plant Requirements in Telephone Offices

The telephone central office power plant must be not only reliable at all times but must meet other exacting requirements. Modern practice has led
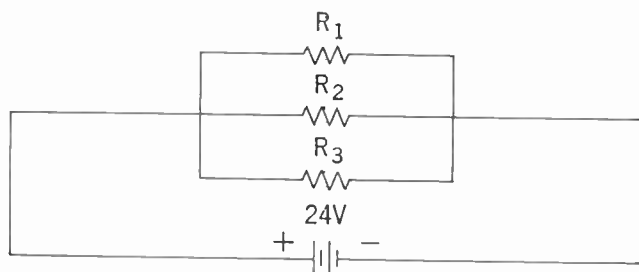
FIGURE 7-4
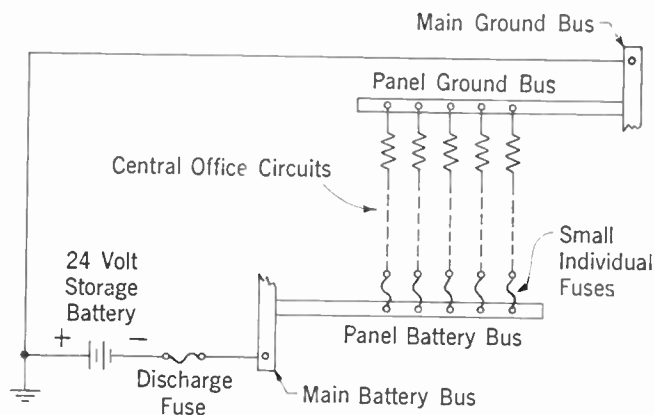
FIG. 7-5 STORAGE BATTERY LOAD CONNECTIONS

to the standardization of a common source of emf for the majority of the talking circuits, as well as for the operation of telephone and telegraph relays and numerous other apparatus units. We thus have a very general use of the standard 24-volt storage battery, with additional smaller batteries used for such services as 48-volt subscriber's transmitter supply on long distance connections, 130-volt supply, both positive and negative, for telegraph repeater operation, and other voltages for telephone repeater and other electron tube operation. The common battery results in a number of plant economies, but, on the other hand, imposes certain exacting electrical requirements. Probably the most essential of these requirements is **low internal resistance.**

In our study of simple electric circuits, we have considered a single source of emf for each individual circuit. But we have learned that any number of resistances may be connected in parallel, as shown by Figure 7-4, and that the current in any single resistance is independent of that in any other resistance provided all resistance branches are connected directly to the terminals of the battery as indicated. This follows naturally from the application of Ohm's Law to a single resistance branch, since the emf impressed on any single branch is the emf of the source and, theoretically, is independent of current flowing through other branches. This assumes, however, that the battery is a perfect source of emf without internal resistance.

Figure 7-5 represents the central office storage battery connected to bus-bars at the fuse panel. The central office circuits are cabled to this fuse panel and receive their battery supply through taps to the small panel busses. Thus hundreds of

[53]

circuits of varying resistance are connected in parallel to a common battery, and we have in practice a circuit arrangement identical to that shown in theory by Figure 7-4, excepting that as indicated in Figure 7-5, fuses for protection against excessive currents due to short-circuit or overload are used, and the positive terminal of the battery is connected to ground. This ground connection stabilizes the potential of all circuits in the central office by short-circuiting their capacitances to ground. It also simplifies the central office wiring and affords circuit protection, but it cannot in any way affect the total current supplied by the battery or the current in any individual circuit that may be connected to the bus-bars.

Returning to Figure 7-4, in which the current in any one resistance branch was seen to be independent of that in any other (provided the source of emf is a perfect one), let us assume, on the contrary, that the battery has an internal resistance R_0 and that the circuit is actually that shown by Figure 7-6. Due to the resistance R_0, the current in one branch is no longer independent of that in other branches. Let us assign values as follows:
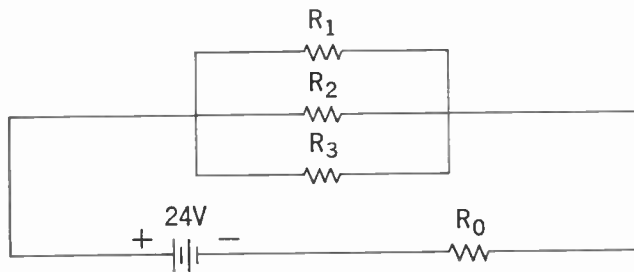
$$R_0 = 2 \text{ ohms}$$

$$R_1 = 5 \text{ ohms}$$

$$R_2 = 4 \text{ ohms}$$

$$R_3 = 3 \text{ ohms}$$

$$E = 24 \text{ volts}$$

If we solve this network, we shall find that the current through R_1 is 1.87 amperes. If we should suddenly open resistances R_2 and R_3, however, it would immediately change to 3.43 amperes. Applying the same principle to Figure 7-5, **unless the central office source of emf has negligible resist-**

FIGURE 7-6

ance including both the internal resistance of the battery and that of the supply leads from the battery to the bus-bars where individual circuit leads are connected, there will be ever-changing current values in the individual circuits. This will result in noise and crosstalk in all talking circuits and unreliable operation of various other telephone apparatus. From this it follows that common battery operation for any number of circuits may be substituted for local or individual batteries only when the common source of emf has negligible internal resistance.

7.5 Rectifiers

Although a complete understanding of the operation of rectifying devices requires a knowledge of alternating currents, which are discussed in later Chapters, a few of the essential characteristics of the more common types of rectifiers may be mentioned here for completeness.

A rectifier is commonly defined as a device for converting alternating electric current to direct electric current. However, there are devices which perform this function which are not normally referred to as rectifiers. For example, an alternating-current motor driving a direct-current generator is referred to as a motor-generator set. If the motor and generator of such a set are combined in one housing with a single rotor, the machine is referred to as a rotary converter. In either case, however, electrical energy is first converted to mechanical energy and this in turn is converted to a different type of electrical energy. Accordingly, rectifiers may be somewhat more precisely defined as devices for converting a-c energy to d-c energy directly or without an intervening step.

All rectifying devices depend for their operation upon the characteristic of permitting electric current to flow through them freely in one direction only. They include a variety of vacuum and gas filled tubes such as the older mercury arc tube, the newer mercury-vapor tube, and the Tungar tubes, as well as nearly all other types of electron tubes when properly connected. There are also the metallic junction types of rectifiers which are represented in their order of development by the copper-oxide rectifier, the selenium rectifier, and the silicon power rectifier. The theory of operation of these semi-conductor devices is discussed at some length in Chapter 23.

The Tungar and mercury-vapor tube rectifiers
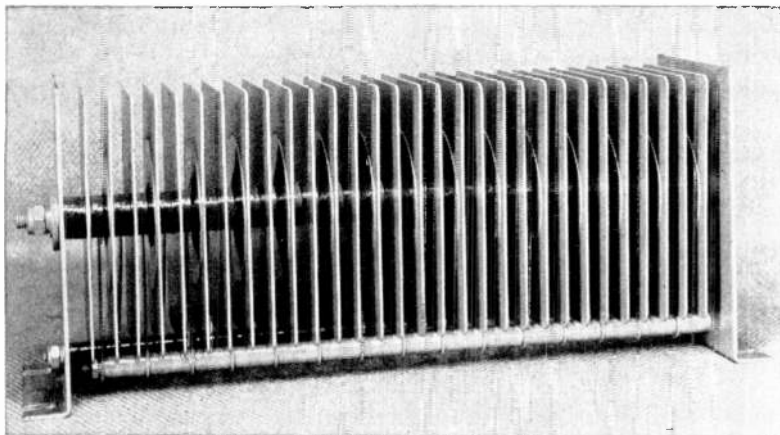
SELENIUM RECTIFIER STACK

depend for their operation upon the emission of electrons from a heated cathode. The basic phenomena involved are the same as characterize all types of electron tubes and these are discussed more fully in a later Chapter. Mercury vapor tubes are used quite extensively in telephone plants for charging small storage batteries and similar purposes.

Metallic rectifiers are also widely used in the telephone plant where they serve an increasing variety of purposes. These uses include the charging of small storage batteries and emergency cells, the furnishing of a direct source of power for small repeater installations, etc. In lower voltage plants (12, 24 and 48-volt) they may be used in place of motor-generator sets in the 100-ampere and 200-ampere sizes.

The copper-oxide rectifier element consists of a copper disc upon which has been formed a layer of copper oxide. This combination offers a low resistance to current flowing from the copper oxide to the copper but a high resistance to current flowing from the copper to the copper oxide.

Thus it becomes a "valve" to pass current in one direction only.

Similarly the selenium rectifier or cell consists of a back plate or electrode, usually of steel or aluminum. One side of this plate is coated with a very thin layer of metallic selenium which in turn is in direct contact with a front or counter electrode of conducting metal. Rectification occurs in the so-called "barrier layer" where the selenium is in contact with the front electrode. Current flows readily from the back electrode to the front electrode but only very little current will flow in the reverse direction.

The recently developed silicon power rectifier is superior in several respects to either of the metallic rectifiers discussed above. Its rectifying element consists of a tiny wafer of silicon, pure to one part in a billion, with a small but controlled amount of impurity diffused into a thin layer of its surface. The junction between this "doped" surface layer and the pure silicon forms the rectifying barrier. For a given size, it can handle much higher power than other metallic junction rectifiers. Thus an element no larger than a pea is capable of furnishing a current of as much as 10 amperes at 100 volts.

A selenium element with capacity comparable to this is much larger in size. Both kinds of rectifier elements can be multipled with like elements in suitable series or parallel connections which, when properly cooled, can furnish large quantities of d-c power. Copper-oxide rectifiers, because of their cost and space requirements, find little employment in power applications although they are extensively used in various modulating circuits, as discussed in Chapter 27, where relatively small current values are involved.

CHAPTER 8

INDUCTANCE AND CAPACITANCE

8.1 Classification of Electric Currents

Thus far we have confined our attention largely to circuits of relatively simple characteristics. We have had a source of direct emf connected to one or more resistances, and have assumed a resultant steady current in each closed branch. We have noted, however, the alternating character of the emf generated by a closed loop revolving in a magnetic field; but we have not attempted to analyze the behavior of such an emf when acting in various types of circuits.

It is desirable at this time that we broaden our studies somewhat to include more general conditions and while nothing that we have learned thus far will be invalidated, it will be necessary for us to study certain additional properties of electric circuits and their effect on the current set up in them by impressed emf's.

Broadly speaking all electric currents may be classified into five groups as follows:

a. The current that results from a constant direct source of emf connected to a resistance network (i.e., the condition assumed in the earlier Chapters for the calculation of direct-current networks through the application of Ohm's and Kirchoff's Laws).

b. The current immediately after opening or closing a circuit, varying its resistance, or in some way interrupting the steady direct current for a short period of time during which the current values readjust themselves before again becoming fixed or steady.

c. Current where the source of emf is alternating, having the simplest, most common and most convenient wave form, viz., the sine wave.

d. Current where the source of emf is alternating, with a definite wave shape other than the sine wave.

e. Alternating current immediately after opening or closing the circuit, or immediately after effecting some other change in circuit conditions.

We can carry this classification one step further

by noting that a, c, and d are **steady state** currents, while b and e are temporary currents, sometimes called **transients.** In practice we are mostly interested in steady state currents in so far as the actual determination of current values is concerned, but under certain conditions the effects of transients are important. Certainly, in a telephone connection, we are concerned with any "clicks" or "scratches" that may be heard in a telephone receiver due to the opening or closing of circuits which are electrically connected to the telephone system. For example, when sending telegraph signals over a telegraph circuit superposed on a telephone circuit, there should be no appreciable "telegraph thump" in the telephone circuit. The successful operation of both telephone and telegraph circuits introduces important considerations having to do with changes in current values.

In fact, we deal in the telephone plant with all five of the circuit conditions mentioned above. Let us consider a long distance line wire not only composited for telegraph service but having a carrier current telegraph channel superposed as well. The resulting current in the wire can best be studied by scrutinizing the behavior of its separate components. When analyzed, the current due to the composited telegraph connection alone is an illustration of two of the classifications, namely a and b. At the instant of "make" or "break" of the key, conditions are as described by b. When the key is closed, i.e., when signals are not being sent, conditions are as described by a. For the carrier channel, we likewise have condition c for a part of the closed key period and condition e for the instants of "make" and "break". For the main talking circuit, we have an application of d when a vowel sound is being transmitted, and an application of e when a consonant sound is being transmitted.

Thus we find in the telephone plant no scarcity of applications for every current classification. It happens, however, that some of these are by no means simple, and for practical telephone work we may limit our study to a thorough analysis of steady state currents only; and to concepts, rather than calculations, of transients in either direct or alternating-current circuits.

[56]
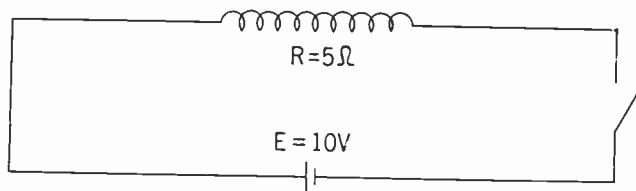
FIG. 8-1 SIMPLE INDUCTIVE CIRCUIT

8.2 Changes in Direct-Current Values

Let us analyze classification *b* (changes in direct-current values), since this will lead us to certain of the new circuit properties that we wish to examine. In Figure 8-1, with the switch open we have a circuit with infinite resistance and zero current; with the switch closed we have, by Ohm's Law, a current—

$$I = \frac{E}{R} = \frac{10}{5} = 2 \text{ amperes}$$

In spite of the apparent promptness with which electricity responds to the operation of any controlling device, we cannot conceive of the current changing from zero to two amperes without going through the range of every intermediate value between zero and two amperes; neither can we conceive of the current building up in the circuit in zero time to the value given by the application of Ohm's Law. If such were the case, the current would have every value from 0 to 2 amperes at the instant of closing the circuit. Reverting to our water analogy with the circulating mechanism in Figure 2-2, when a valve is shut we know there is no flow of water in a long pipe line and when the valve is opened we know that, due to the inertia of the water, a definite time is required for the flow to become a maximum. A current in an electrical circuit cannot be established instantaneously any more than the water flow can be established instantaneously.

Again, if in Figure 8-1 we suddenly open the switch in a dark room while there is a current of two amperes in the circuit, we shall observe a spark at the contacts of the switch. Though the electric current is reducing in value, it continues to flow for an instant after the switch points are no longer in contact, forcing itself through the air, and thereby forming an "electric arc" which gives the illllumination.

We thus have two conditions where the current in a brief interval of time assumes all intermediate values between two amperes and zero, and we may compare these with other less abrupt changes in a circuit. It may be said that an electric circuit "reacts" to such current changes. But this reaction cannot be explained by our previous understanding of either resistance or emf. The circuit has other properties which are latent when the current is a steady unidirectional one but which are immediately brought into play when the current attempts to change its value. There are two such additional properties, namely, **"inductance" and "capacitance". Inductance tends to give the circuit something that is analogous to inertia in a mechanical device, and capacitance something analogous to elasticity.**

8.3 Inductance

When an emf is connected to a circuit, the conditions are somewhat analogous to those obtaining when a locomotive starts a train. The locomotive exerts considerable force which, in the circuit, corresponds to the impressed emf. A part of this force is used in overcoming resisting forces such as the friction of the moving wheels, the grade of the track, and others that apply to the train as a definite resistance to its motion at all times. The second part of the force is used in setting the train in motion, i.e., accelerating the heavy inert body. As soon as the train is accelerated to full speed, the entire force applied is available for overcoming the resistance alone. Likewise in the electric circuit, for any given emf, the current does not instantaneously establish itself to that value which represents the effect of the full voltage overcoming the resistance.

We have learned that there is a magnetic field about every current-carrying conductor, and when a conductor is wound into a coil or is in the presence of iron, the magnetic field is intensified. The magnetic field cannot be established instantaneously any more than the train can be instantly changed from its state of rest to that of full speed. What actually happens in the case of the electric circuit is that the emf endeavors to start a current; the current in turn must establish a magnetic field; this field reacts upon the circuit in a manner similar to that in which the counter-emf generated by a motor opposes the applied voltage, and for an instant a part of the emf that is connected to the circuit must be used in overcoming these reactions. The current, therefore, increases gradually and as it does so, the magnetic field becomes more nearly established and the reaction becomes less pronounced, until finally the entire
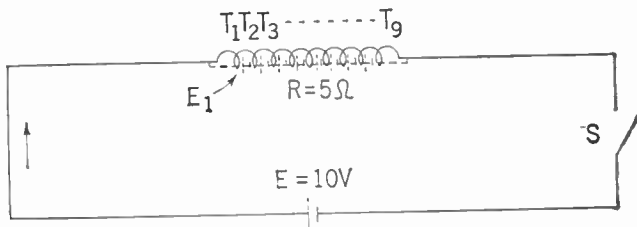
FIGURE 8-2

emf is applied to overcoming the resistance of the circuit alone, thereby sustaining the established current at a value determined by Ohm's Law.

This may be more clearly understood by referring to the circuit shown in Figure 8-2. Immediately after the switch has been closed, the emf E endeavors to establish a current in the circuit equal in value to E/R, or two amperes. But the current, as has been stated, must go through every intermediate value from zero to two amperes. By directing our attention to only one turn of the coil, for example, T_1, we can imagine the current building up and in consequence establishing lines of magnetic induction around this single turn which will, however, cut every other turn of the coil. This action will set up in the other turns an induced emf tending to establish a current in the opposite direction. The two currents are in one and the same circuit and the induced current is opposed to the current established by virtue of the battery emf.

Figure 8-3 represents graphically the current in this circuit. With the switch open, the current is zero. When it is closed (or when sufficiently near the contacts for the emf to break down the insulation of the narrow separation of air, since the current starts to flow before actual contact is made), the 10-volt battery will attempt to establish a current of two amperes in accordance with Ohm's Law. But the current cannot be completely established until after an interval of time represented by t_2; and at the start, it cannot be increasing at a rate greater than that which would induce a back emf of 10 volts, because if it did so, the induced emf would be equal to the applied emf and since they oppose each other, there would be no current whatsoever. As would be expected, however, the maximum rate of increase of the current occurs at the instant the switch is closed.

Now let us consider the conditions at some intermediate time between the closing of the switch and t_2. If, from the value represented by point P, the current increased at a rate that continued

without changing, the line PM would represent the trend of current values that would follow. But with the current increasing at this rate, the lines of magnetic induction are cutting other turns of wire and inducing an emf which we might represent in Figure 8-2 as a second battery E_1. The voltage of E_1 must be of a value necessary to establish a current equal to two amperes minus the current which has been already established at the point P. This follows from the earlier explanation regarding the directional property of an induced emf. If the battery voltage E acted alone, the current value would be E/R or two amperes. Since the actual current flowing is less than two amperes, the difference between the actual current and two amperes may be regarded as due to a current flowing in a direction opposite to that of the two amperes set up by the battery. This current is established by the induced emf and we may designate it as an **induced current** to distinguish it from the two-ampere current which the supply voltage tends to set up. The actual current in the circuit at any instant, then, is the numerical difference between the two-ampere battery current and the induced current.

If we now assume for the sake of reasoning that the induced voltage E_1 remains unchanged, the resulting induced current will oppose the battery current, and the net amount of current flow will remain at the value P. We know, however, that the current which will eventually flow is two amperes, and furthermore, if the current becomes constant at a value P, no lines of magnetic induction are in motion; hence, there is no induced voltage and consequently no induced current. But with no induced current, the battery will set up
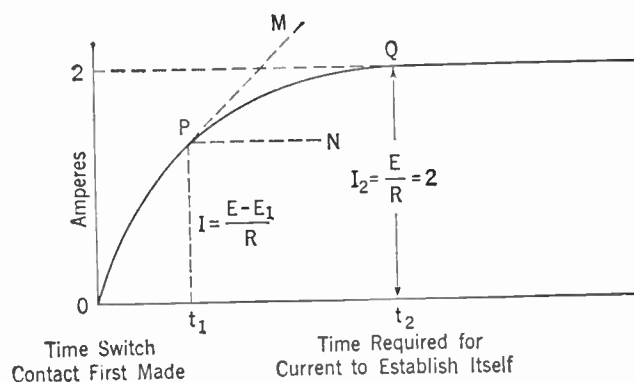
FIG. 8-3 BUILD-UP OF CURRENT IN INDUCTIVE CIRCUIT

[58]

two amperes; therefore our assumption that the induced voltage E_1 remains constant, keeping the current down to a value such as that represented by the line PN is false. On the other hand, it is clear that the induced voltage E_1 cannot become zero until the current becomes two amperes, though it does continue to decrease in value, since we know that a current is always accompanied by a magnetic field which must change if the current changes, and the result of such a change is an induced voltage. From this we conclude that there must be a compromise trend for the curve of current as it establishes itself, somewhere between the two extremes. This compromise is that shown by the curve PQ which is tangent to but bending away from PM. The current is neither maintaining the same rate of change as it approaches the value fixed by Ohm's Law nor does it cease entirely its increase in value before it reaches two amperes. This is true because although the induced emf that would stop the change in current is gradually becoming less in value, the IR drop is becoming greater, and the sum of these two must always equal the impressed voltage in accordance with Kirchoff's second law. Thus we see from the curve in Figure 8-3 the "choking" effect of an inductively wound coil to increases in current value.

The case of a decreasing current value, and the emf induced at the time of opening a circuit, is of course another application of the same theory, but the effects are different in their practical aspects. Because this emf is induced as a result of a decreasing current instead of an increasing one, it aids rather than opposes the existing emf. Moreover, the current change is a very rapid one because the opening of the switch tends to change the resistance of the circuit from a definite value to infinity with great suddenness. As a result, the induced emf may become much greater than the applied emf besides being additive to it, whereas in the closed circuit it can never be greater than the applied emf. This total emf of the opening circuit tends to force an arc across the switch contacts, which is much more evident than the arc at the time of closing the switch because the voltage is so much greater. Here we have again the analogy to inertia where we attempt to suddenly stop a moving body, whereas before the analogy covered starting a body from a state of rest.

Briefly, Ohm's Law holds at all times, but the property of inductance in a circuit will cause the establishment of an emf opposing that applied to the circuit in the case of an increasing current, and aiding the applied emf in the case of a decreasing current. The value of this induced emf is not necessarily a fixed one; it varies, and either in the case of a current establishing itself, or in the case of a current decreasing, eventually becomes zero. The magnitude or influence of the induced emf as a reactive effect is determined by two factors:

a. The first is a property of the circuit having to do with the number of inductive turns, whether or not each coil has a magnetic core and if magnetic, the permeability of the iron, etc.
b. The second is the rate of change of current. This employs the foregoing property of the circuit as a tool or facility for creating the induced emf.

The **property of the circuit** which we have called inductance is represented by the symbol L and is measured in a unit called the **henry**. **The unit value of the henry is defined as the inductance of a circuit that will cause an induced emf of one volt to be set up in the circuit when the current is changing at the rate of one ampere per second.** From which we may write—

$$E_1 = \frac{LI}{t} \qquad (8:1)$$

where E_1 is the symbol for induced emf and L represents inductance in henrys.

Since L depends upon a property of the circuit which has to do with conductors cutting lines of magnetic induction, it can be defined in other terms. In Chapter 6 we learned that an emf induced in a conductor cutting through a magnetic field was proportional to the number of lines of magnetic force cut per second. Then we were considering a moving conductor and a stationary field. Now we are considering the stationary conductors of a coil and a moving field in which the magnetic lines threading through the coil will cut each of its turns as they build up or decrease. The voltage induced remains proportional to the number of lines cut per second, or we may write—

$$E_1 = k_1 \frac{\phi N}{t} \qquad (8:2)$$

where ϕ is the total flux through the coil, N is the number of turns, and k_1 is a constant depending
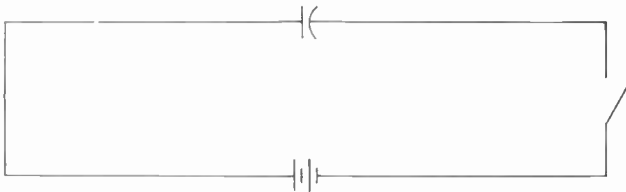
FIG. 8-4 SIMPLE CAPACITIVE CIRCUIT

on the units employed. Then from equations (8:1) and (8:2) we have

$$\frac{LI}{t} = k_1 \frac{\phi N}{t}$$

or

$$LI = k_1 \phi N \qquad (8:3)$$

Now referring to Chapter 4 we find the following two equations—

$$H = k \frac{NI}{l} \qquad (4:3)$$

and

$$\frac{\phi}{A} = H\mu \qquad (4:5)$$

Substituting H from equation (4:3) in equation (4:5) we get

$$\phi = k\mu A \times \frac{NI}{l}$$

and substituting this in equation (8:3) we have—

$$LI = k\,k_1 \frac{\mu A N^2 I}{l}$$

or

$$L = k\,k_1 \frac{N^2 \mu A}{l} \qquad (8:4)$$

Here, if L is in henrys, N is the number of turns of the coil, μ is the permeability (in the cgs electromagnetic system of units), l is the length of the coil core in centimeters and A is the area of the core in square centimeters, the combined value of the constants k, k_1 will be 1.26×10^{-8}, or we may write—

$$L = \frac{1.26\,N^2 \mu A}{100,000,000 \times l} \qquad (8:5)$$

Note: This equation may be used to calculate the inductance of a coil if all of the constants involved are accurately known and there is no flux leakage. In practice, it is usually easier to measure the inductance.

If it is desired to find the total inductance of a circuit having several coils in series, the inductances should be added in the same way that resistances in series are added. Similarly, parallel inductances are calculated by the same equations as are parallel resistances. For example, see equation (3:1) and substitute L, L_1, and L_2 for R, R_1 and R_2, respectively, etc.

This property of a circuit which creates an emf from a change of current values when the reaction effects are wholly within the circuit itself is called self-inductance to distinguish it from the relation permitting electromagnetic induction between coils or conductors of separate circuits. This latter property of the two circuits taken jointly is called mutual inductance. It is discussed in a later Chapter.

8.4 Capacitance

There remains that property of the circuit that we have called **capacitance**, which gives it something analogous to elasticity. While a storage battery stores electricity as another form of energy, in a smaller way an electric "condenser" or **capacitor** stores electricity in its natural state.

As a container, a capacitor is hardly analogous to a vessel that may be filled with water, but more nearly to a closed tank filled with compressed air. The quantity of air, since air is elastic, depends upon the pressure as well as the size or capacity of the tank. If a capacitor is connected to a direct source of emf through a switch as shown in Figure 8-4, and the switch is suddenly closed, there will be a rush of current in the circuit. This will charge the capacitor to a potential equal to that of the battery, but the current will decrease rapidly and become zero when the capacitor is fully charged.

The insulated conductors of every circuit have to a greater or less degree this property of capacitance. A certain quantity of electricity, rep-
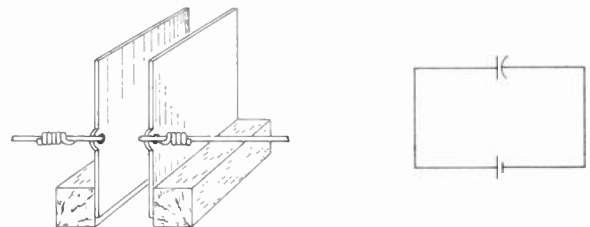
FIG. 8-5 ELEMENTARY CAPACITOR AND CONVENTION

resenting a certain quantity of energy, is accordingly delivered to a circuit before the actual transfer or transmission of energy from a sending device to a receiving device takes place. The capacitance of two parallel open wires, or a pair of cable conductors of any considerable length, is appreciable in practice.

TABLE IV

Dielectric Power of Various Insulating Materials

Values are only approximate and are given for value of K in equation (8:8) rather than compared to air as unity.

SUBSTANCE	K IN EQUATION (8:8)
Air	$.09 \times 10^{-6}$
Alsimag A196	$.51 \times 10^{-6}$
Bakelite (paper-base)	$.34$ to $.50 \times 10^{-6}$
Bakelite (mica-filled)	$.49 \times 10^{-6}$
Celluloid	$.36$ to 1.4×10^{-6}
Cellulose Acetate	$.54$ to $.72 \times 10^{-6}$
Fiber	$.45$ to $.64 \times 10^{-6}$
Formica	$.41$ to $.44 \times 10^{-6}$
Glass (window)	$.68$ to $.72 \times 10^{-6}$
Glass (photographic)	$.68 \times 10^{-6}$
Glass (Pyrex)	$.38$ to $.44 \times 10^{-6}$
Lucite	$.23$ to $.27 \times 10^{-6}$
Mica	$.23$ to $.72 \times 10^{-6}$
Mica (clear India)	$.58$ to $.68 \times 10^{-6}$
Mycalex	$.67 \times 10^{-6}$
Paper	$.18$ to $.23 \times 10^{-6}$
Polyethylene	$.21$ to $.22 \times 10^{-6}$
Polystyrene	$.22$ to $.26 \times 10^{-6}$
Porcelain	$.56$ to $.68 \times 10^{-6}$
Rubber (hard)	$.18$ to $.32 \times 10^{-6}$
Steatite (low-loss)	$.40 \times 10^{-6}$
Wood (dry Oak)	$.23$ to $.61 \times 10^{-6}$

The quantity of electricity stored by a capacitor depends upon the capacitance and the electromotive force impressed across its terminals. The following equation expresses the exact relation:

$$Q = EC \qquad (8:6)$$

where Q is the quantity of electricity in coulombs, E is the impressed emf in volts, and C is the capacitance in **farads**. The farad is a very large unit and is seldom used in practice. The microfarad (from "micro", meaning one-millionth) is the practical unit more commonly used; and with C expressed in these units, equation (8:6) becomes—

$$Q = \frac{EC}{1,000,000} \qquad (8:7)$$

Figure 8-5 illustrates a capacitor in its simplest form together with the convention used for a capacitor connected to a battery. Two wires are connected to two parallel metal plates having a definite separation as shown. This is called an **air capacitor** because air is the **dielectric** medium between the plates. The capacitance of such a capacitor is directly proportional to the area of the plates, and inversely proportional to their separation. At the instant a battery is connected to its terminals, there is a rush of electricity which charges the plates to the potential of the battery, but as the plates become fully charged, the current in the connecting conductors becomes zero. Were we to insert a sensitive high resistance galvanometer in series with the battery, we would observe an instantaneous "kick" of the needle when the connection is made, but the needle would return and come to rest at zero. If the capacitance of the capacitor were increased, the kick would become more noticeable. If now the battery were disconnected and the capacitor short-circuited through the galvanometer, there would be a kick of the needle in the opposite direction. This would result from the quantity of electricity, which had been stored in the capacitor, establishing an instantaneous current in the opposite direction and discharging the capacitor through the winding of the galvanometer.

In addition to the size of its plates and their separation, the capacitance of a capacitor depends upon the insulating medium between the plates. For example, if mica is inserted between the plates of an air capacitor its capacitance is increased about five times. Insulators, in addition to being classified in the order of their insulating properties, may be classified in the order of their "dielectric powers," or "specific inductive capacities," i.e., their ability to increase the capacitance over that of an air capacitor. Such a classification is given in brief in Table IV.

The equation for the capacitance value of a two-plate capacitor is—

$$C = K \frac{A}{d} \qquad (8:8)$$

where C is capacitance in microfarads, K is the constant taken from Table IV, A is inside area of

FIGURE 8-6

one plate in square centimeters, and d is separation of the plates in centimeters. There are similar equations for calculating the capacitance per unit length of parallel open wire conductors or cable conductors. These may be found in various handbooks, but for telephone and telegraph work, tables giving measured values, which vary for each class of open wire or cable pairs, are preferable and are usually available.

An inspection of equation (8:8) will show that if two identical capacitors are connected in parallel as shown by Figure 8-6, the effect is that of doubling the plate area of a single capacitor, and therefore doubling the capacitance. On the other hand, if two identical capacitors are connected in series as shown by Figure 8-7, the middle or common plates have a neutral potential and the effect is that of doubling the thickness of the dielectric of a single capacitor, which cuts the capacitance

FIGURE 8-7

in half. It follows that capacitors in parallel and series act inversely to resistors in parallel and series. This may be stated in a single rule covering all conditions—

Capacitors in parallel should be added to find the total capacitance in the same way that resistors in series should be added to find the total resistance; and the reciprocal of the sum of the reciprocals must be taken to find the total capacitance of capacitors in series in the same way that the reciprocal of the sum of the reciprocals must be taken to find the total resistance of resistors in parallel.

This rule may be expressed by two simple equations:

For several parallel capacitors—

$$C = C_1 + C_2 + C_3 \text{ etc.} \qquad (8:9)$$

For several series capacitors—

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \text{ etc.} \qquad (8:10)$$

Or, for the simple case of only two series capacitors, equation (8:10) may be expressed as follows:

$$C = \frac{C_1 \times C_2}{C_1 + C_2} \qquad (8:11)$$

Note: Equation (8:9) may be compared with equation (3:1) and equation (8:11) may be compared with equation (3:4).

8.5 Time Constant

It was noted in Article 8.3 that when a battery is connected across an inductance, the current builds up gradually to the maximum value determined by the resistance of the circuit. Conversely, when a battery is connected across a capacitor, there is a surge of current into the capacitor, which decreases gradually toward zero as the capacitor charges up to the value of the impressed voltage. Thus, in the series $R\text{-}C$ circuit diagrammed in Figure 8-8, the curve I illustrates how the current will change with time after the switch is closed, while the corresponding change in the voltage drop across the capacitor is shown by the curve E. The instantaneous current value when the switch is closed is determined solely by the value of the resistance R. The total voltage drop is then across R and the drop across C is zero. As the capacitor begins to charge, however, the voltage drop across C gradually builds up, the

FIGURE 8-8

[62]
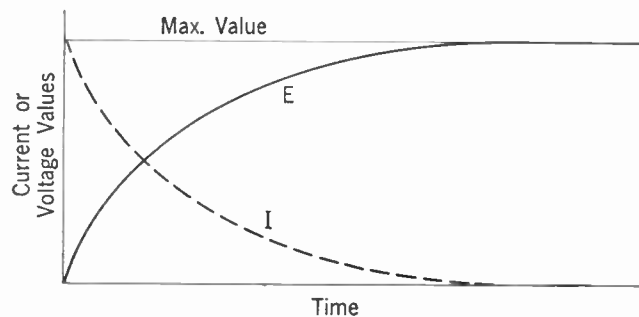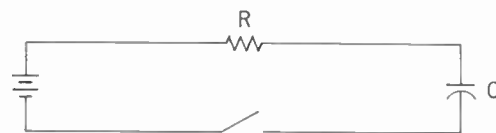
value of current decreases, and the drop across R decreases correspondingly. When the capacitor reaches full charge and the current has fallen to a negligible value, the total voltage drop is across the capacitor. At all times, of course, the sum of the voltage drops across R and C must be equal and opposite to the impressed voltage.

It is an interesting and important fact that the product of resistance and capacitance has the dimensions of time. Actually, resistance in ohms multiplied by capacitance in farads equals time in seconds. In a series resistance-capacitance circuit, the product RC is known as the **time constant** of the circuit. It is the time required to charge the capacitor to 63% of its final voltage. This may be verified by expressing the charging curve of Figure 8-8 mathematically as an "exponential function" of time—

$$E_i = E_{max} \left(1 - \varepsilon^{-\frac{t}{RC}} \right) \qquad (8:12)$$

when E_i is the instantaneous voltage across the capacitor at any time t, E_{max} is the final voltage and ε is the base of Naperian or natural logarithms, $2.718+$. When t is equal to RC, the term in the parenthesis reduces to

$$1 - \varepsilon^{-1} = 1 - \frac{1}{\varepsilon} = 1 - \frac{1}{2.718} = .63$$

A similar expression may be written for the exponential curve of Figure 8-3 illustrating current build-up in a series circuit containing resistance and inductance. Here the time constant is equal to L/R. Again it is the time required for the current to build up to 63% of its final value.

These exact relationships between electrical units and time are invaluable to the designer of modern electronic devices. Because of the precision with which resistors, capacitors and inductors may now be built, they enable him readily to control the timing of current pulses to an accuracy in the order of a few milli-microseconds. This is essential to the satisfactory operation of the high-speed devices employed in television, radar, electronic switching systems, etc. In practical application, the designer assumes that currents and voltages in R-C or R-L circuits reach their final value in a time equal to five times the time constant (i.e., $5RC$ or $5L/R$). This is usually a sufficiently accurate approximation since exact calculation shows that in this time ($5 \times$ time constant) the current or voltage will have reached $99.33+\%$ of its ultimate value.

CHAPTER 9

TELEPHONE PRINCIPLES AND BASIC APPARATUS

9.1 Sound

The telephone accomplishes the electrical transmission of speech by employing the mechanical energy of the speaker's voice to produce electric energy having similar characteristics, and in turn converting this electric energy into sound waves having similar characteristics at the listener's station. To understand its principle of operation we may well consider the nature of **sound**.
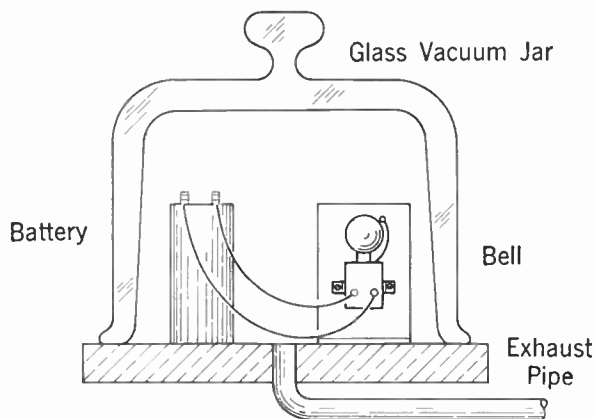
FIGURE 9-1

Sound in the scientific sense has two distinct meanings. To the psychologist it means a sensation, to the physicist it means an atmospheric disturbance or a stimulus whereby a sensation is produced in the human ear. In other words, it is a form of wave motion produced by some vibrating body such as a bell, tuning fork, the human vocal cords, or similar objects capable of producing rapid to-and-fro or vibratory motion.

Everyone is familiar with the series of waves that emanate from a stone cast upon the still water of a lake or pond. This is one of many forms of wave motion, and in a manner similar to that in which the stone coming in contact with the water establishes radiating rings formed by circular wave crests alternating with wave troughs, there emanate from a source of sound alternate condensations and rarefactions of the air. Instead of being rings on a single plane or

surface, however, they are a series of concentric spheres expanding at a definite rate of travel. This rate of travel (or the velocity of the sound wave) is approximately 1,075 feet per second but varies to some extent with altitude and atmospheric conditions. The velocity of sound is very low as compared with the velocity of light, heat or radio waves, which are also a form of wave motion. We thus see a flash of lightning before we hear a clap of thunder or see the smoke dispelled from the muzzle of a gun before we hear the gun's report.

Unlike light, heat or electromagnetic wave transmission, sound is an atmospheric disturbance. If, as shown in Figure 9-1, a vibrating bell is placed under an inverted glass bowl resting on a plate that has an outlet to which an exhaust pump is connected, it may be heard almost as distinctly as though there were no glass container. But if the air is exhausted until there is a vacuum about the bell, no sound can be heard; yet the bell may be seen vibrating as clearly as before the glass container was exhausted. We thus learn that there must be a physical medium, usually atmospheric, for the transmission of sound.

If the sound's source is a vibrating mechanism in simple form, such as a simple to-and-fro motion of a tuning fork, and is sustained for a definite interval of time, the wave motion is said to be "simple harmonic". (A simple harmonic wave may be represented by the sine curve already dis-

Musical Note

Noise
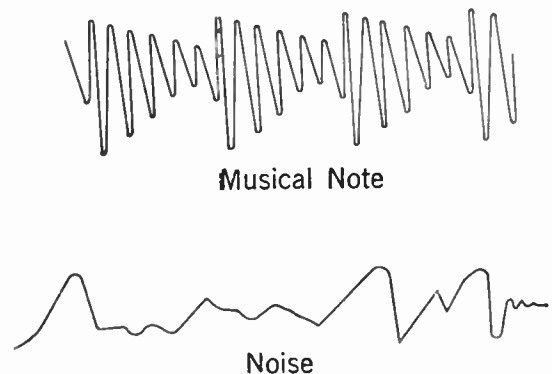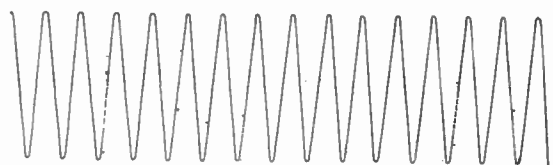
FIGURE 9-2

Simple Sound
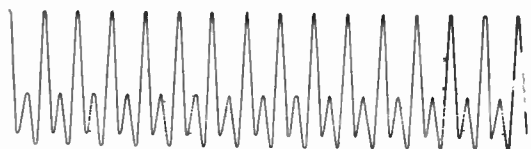
o͞o as in Loose
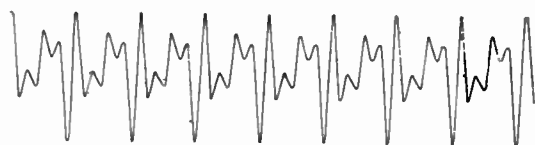
ō as in Low

FIGURE 9-3

cussed in Article 6.2.) On the other hand, if the source consists of a complex mechanical motion or an object vibrating by parts as well as in its entirety, the wave is complex, or a fundamental sine wave with harmonics, the latter giving it "quality". A sine wave without harmonics is called a pure wave.

The sound sensation produced by a series of successive waves identical in form is called a tone, and if each wave is complex, it is a tone having timbre or quality, but if simple or a sine wave, it is a pure tone.

A vibrating mechanism giving a pure tone is said to establish a tone of low pitch if it is vibrating slowly, but if vibrating rapidly, it establishes a tone of high pitch. The lowest pitch which is audible to the normal ear lies somewhere in the octave between 16 and 32 vibrations per second. On the other hand, the normal ear has an upper limit of audibility lying somewhere in the octave between 16,000 and 32,000 vibrations per second. These two octaves are the extreme limits of the scale of audibility.

Audible sound is thus conveniently defined as a disturbance in the atmosphere whereby a form of wave motion is propagated from some source at a velocity of about 1,075 feet per second, the

transmission being accomplished by alternate condensations and rarefactions of the atmosphere in cycles having a fundamental frequency ranging somewhere between 16 per second and 32,000 per second.

The waves superposed on the fundamental, which we have called harmonics, are present in most distinctive sounds, and particularly in the human voice. They permit us to distinguish notes of different musical instruments when sounded at the same pitch. They also establish subtle differences in the voice which may indicate anger or joy, or permit us to distinguish the voice of one person from that of another. Figure 9-2 illustrates wave forms for different kinds of sound and, similarly, Figure 9-3 shows the predominating wave shapes of certain spoken vowels.

Fortunately, in telephone transmission, which is essentially a problem of conveying "intelligibility" from the speaker to the listener, we are not seriously concerned with sounds having either fundamental or harmonic frequencies that extend throughout the entire scale of audibility. The sound frequencies which play the most important part in rendering the spoken words of ordinary conversation intelligible are the band of frequencies within the audible scale ranging from approximately 200 to 3,500 cycles per second.

9.2 The Simple Telephone Circuit

The original telephone, as invented by Bell in 1876, consisted of a ruggedly constructed telephone receiver, which at that time served as both transmitter and receiver. The telephone circuit in its simplest form consisted of two wires terminated at each end by such an instrument but without transmitter or battery and without signaling features. Figure 9-4 shows such a circuit.

At the speaker's station, the sound waves of the voice strike the metal diaphragm of the telephone receiver, and the alternate condensations and rarefactions of the air on one side of the diaphragm establish in it a sympathetic vibration. Behind the diaphragm is a permanent bar magnet and the

FIG. 9-4 ELEMENTARY TELEPHONE CIRCUIT
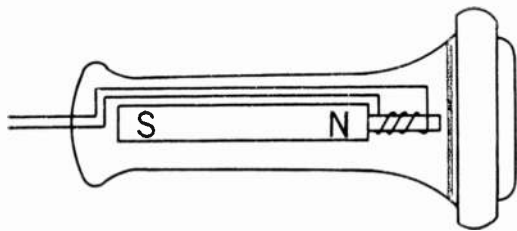
FIG. 9-5 BAR MAGNET RECEIVER

lines of induction leaving the magnet are crowded in the vicinity of the metal diaphragm. The vibration of this diaphragm causes a corresponding change in the number of lines that thread through the receiver winding, resulting in the turns of the winding being cut by these building up and collapsing lines. This establishes a varying electric voltage and current in the winding of the telephone receiver, having wave characteristics similar to the characteristics of the sound wave. This current passes over the connecting wires and through the receiver winding at the distant end. There it alternately strengthens and weakens the magnetic field of the permanent magnet, thereby lessening and increasing the pull upon the receiving diaphragm, which causes it to vibrate in unison with the diaphragm at the transmitting end, although with less amplitude. This vibrating diaphragm reproduces the original sound, conveying intelligibility to the listener at the receiving end.

9.3 The Telephone Receiver

The earliest forms of telephone receivers were made with a permanent bar magnet as shown in Figure 9-5. The efficiency of the receiver was later greatly increased by the use of a horseshoe magnet as shown in Figure 9-6. This permits the lines of magnetic force to pass in a much shorter path from one magnetic pole to the other through the iron diaphragm. The principle of operation
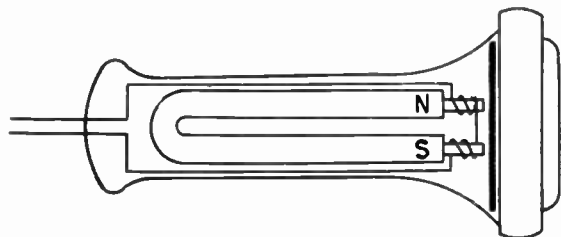
FIG. 9-6 HORSESHOE-MAGNET RECEIVER

of receivers currently in use in the telephone plant does not differ fundamentally from that of the early types although the receivers themselves are generally quite dissimilar in physical appearance.

Figure 9-7 is a cross-sectional drawing of a bipolar receiving unit which is in extensive use in the Bell System. This receiver employs in its construction no less than three of the comparatively new magnetic alloys that were mentioned in Article 4.1. It also differs notably from earlier types in the extent to which the motion of the diaphragm, which is made of vanadium permendur, is affected by "acoustic controls". One acoustic control is directly behind the diaphragm,
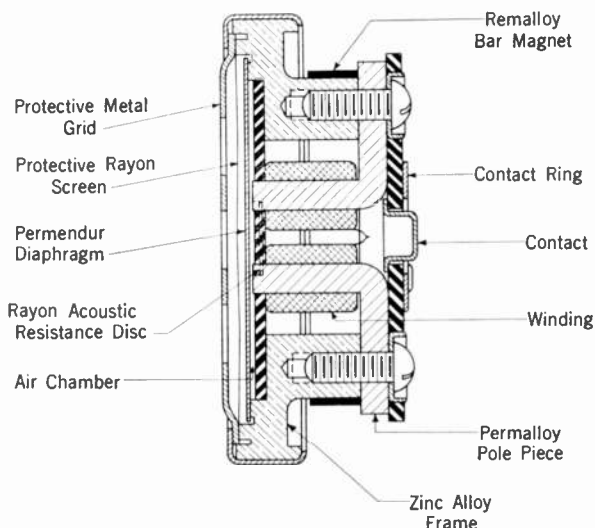
FIG. 9-7 CROSS-SECTION OF BIPOLAR RECEIVER UNIT

and the other is between the diaphragm and the inner surface of the receiver cap when the receiver unit is mounted in the telephone instrument. The former control consists of an air chamber with an outlet to the back of the receiver unit through a small hole covered with a rayon disc. The latter consists of an air chamber which opens into the air through six holes in the receiver cap. These air chambers are designed to have "acoustic impedances" which match the "electrical impedances" of the receiver and improve its overall efficiency appreciably. The diaphragm rests on a ring-shaped ridge and is held in place by the pull of the magnet. In this way variations in receiver efficiency at different frequencies are held to close limits. The two permalloy pole pieces are welded to a pair of very strong remalloy bar magnets, and the assembly is

RING-ARMATURE RECEIVER
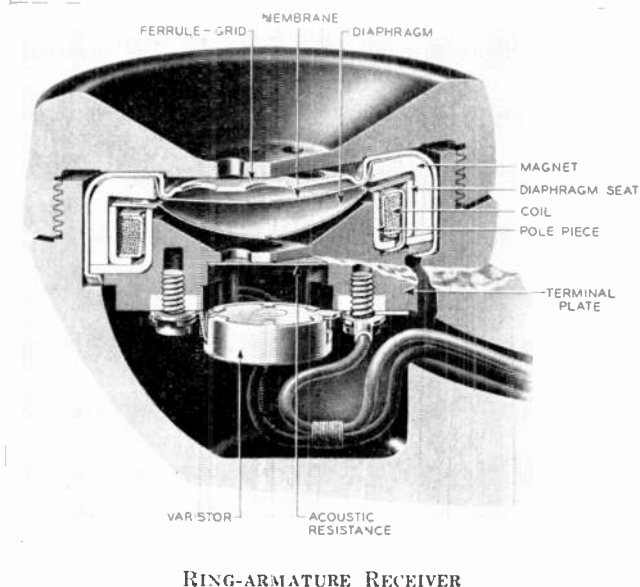
fastened to a zinc alloy frame. The whole unit is held together by an aluminum ferrule on the back of which are mounted two silver plated contacts for the electrical connections.

A more recent design of receiver, which is currently standard for new installations, is shown in cross-section in the accompanying photograph. This is known as a **ring-armature** receiver. Differing radically in detail design from the bipolar types, it is substantially more efficient chiefly because the dome-shaped diaphragm is much lighter than that of the bipolar receiver. The diaphragm dome is made of lightweight plastic-impregnated cloth cemented at its outer periphery to a flat ring-shaped armature of vanadium permendur. The outer edge of the armature ring rests on a diaphragm seat of non-magnetic material and its inner edge extends into the air-gap between a pole piece of 45% permalloy and the permanent magnet which is made of remalloy. The entire diaphragm structure is thus driven like a piston under the influence of the magnetic fields existing in the air-gap across the inner edge of the armature ring.

Since telephone receivers are equipped with permanent magnets, it is of course important that the magnetism should not be impaired by jarring or other abuse. The permanent magnet is important not only because it increases the amplitude of vibration of the diaphragm when the voice current is flowing through the windings, but also because it prevents the diaphragm vibrating at twice the voice frequency. This principle is illus-

trated in Figure 9-8. When a piece of soft iron is held near an electromagnet, it is attracted by the magnet regardless of the direction of the current in the windings. Thus, an alternating current in a winding on a soft iron core will assert an attraction during each half cycle, which in the case of the receiver diaphragm will establish a vibration with a frequency twice that of the current. If, on the other hand, a permanent magnet is used, the alternating current establishes a vibration of the same frequency as the current by merely increasing or lessening the pull already exerted on the diaphragm.

9.4 The Telephone Transmitter

Although the principle of Bell's original telephone applies to the present day telephone receiver, it was appreciated in the early stages of telephone development that the electrical energy generated by a diaphragm vibrating in a comparatively weak magnetic field was insufficient for the transmission of speech over any considerable distance. The energy could, of course, be increased by using stronger magnets, louder sounds, and the best possible diaphragms, but even with any ideal telephone receiver that might be perfected, voice transmission would be limited to comparatively short distances. One year after the invention of the original telephone, the Blake transmitter was introduced. It worked on the principle of a diaphragm varying the strength of an already established electric current, instead of generating electric energy by means of electromagnetic induction. By this means it was possible to establish an electric current with an energy
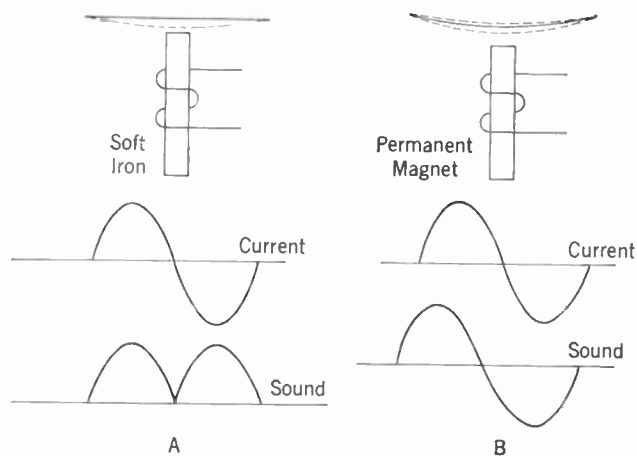
FIGURE 9-8
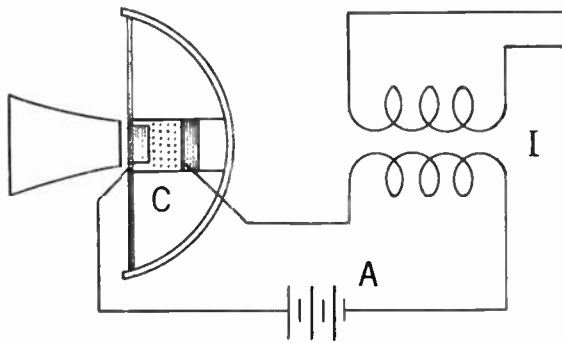
FIG. 9-9 PRINCIPLE OF THE TELEPHONE TRANSMITTER

value much greater than that conveyed to the instrument by a feeble sound wave. The battery in this case was the chief source of energy and the vibration of the diaphragm acted as a means for regulating or modulating this energy supply, rather than as a generating device.

The principle of the transmitter is illustrated by Figure 9-9. Battery A establishes a direct current in a local circuit consisting of the primary winding of an induction coil I, and a cup of carbon granules C. One side of this cup rests against a small carbon disc rigidly connected to the transmitter diaphragm. The vibrating transmitter diaphragm varies the pressure on the carbon granules, which causes the resistance of the electric circuit through the carbon granules to vary correspondingly, thereby causing fluctuations in the value of the direct current maintained in the circuit by the battery. These fluctuations, though represented by varying direct-current values instead of by an alternating current, as in the case of the telephone circuit in Figure 9-4, establish an alternating emf in the secondary winding of the induction coil. This, in turn, sets up an alternating current through the local receiver, over the line, and through the distant receiver. The operation of the distant receiver is the same as has been explained in connection with Figure 9-4.
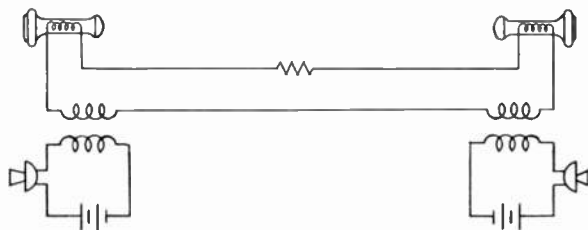
FIG. 9-10 TELEPHONE CIRCUIT WITH LOCAL BATTERY TRANSMITTERS

Figure 9-10 shows transmitters used at the ends of a simple telephone circuit. When the magnetic field is established by the fluctuating current through the primary of the induction coil, an alternating current is induced in the secondary of the coil. This current flows through the receiver at the same end of the circuit, giving "side-tone" to the receiver at the home station. It is also transmitted to the distant station, operating the receiver at that point.

Figure 9-11 shows in cross-section a transmitter unit that is standard for subscribers' telephone sets. This transmitter is of the "direct action" type; that is, the movable element attached to the diaphragm which actuates the granular carbon is an electrode, and serves the dual pur-
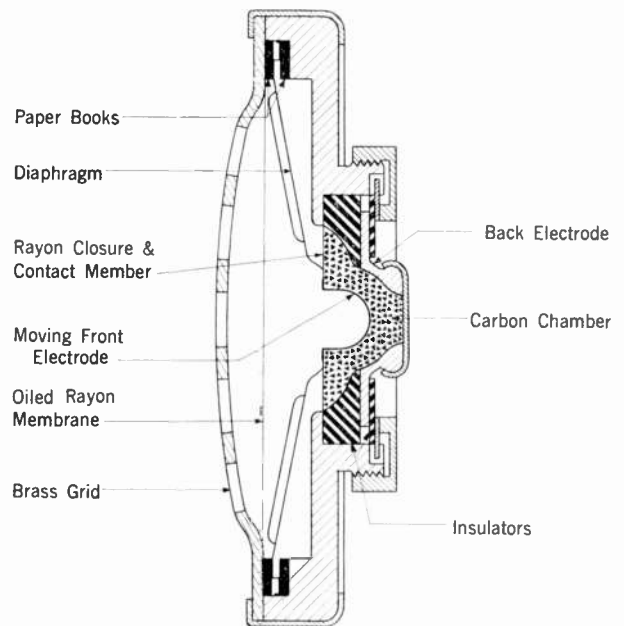
FIG. 9-11 CROSS-SECTION OF TRANSMITTER UNIT

pose of contact and pressure surface. As the drawing shows, this dome-shaped electrode is attached to the center of a conical diaphragm, and forms the front center surface of the bell-shaped carbon chamber.

The diaphragm is made of aluminum alloy with radial ridges to increase stiffness. Paper books, which consist of a number of thin impregnated paper rings, support the diaphragm at its edges on both sides. The carbon chamber is closed on the front side by a rayon membrane clamped under the flange of the diaphragm electrode. A

[68]

light, spoked copper contact member, clamped under the diaphragm electrode, is the means of providing a flexible connection between this front electrode and the supporting metal frame. The fixed back electrode is held in place in the frame by a threaded ring and is insulated by a phenol fibre washer and a ceramic insulator which also forms one of the surfaces of the carbon chamber. The active surfaces of both electrodes are gold plated. A brass plate which is perforated with large holes protects the vibrating parts against mechanical injury. Moisture is kept out of the working parts by an oiled rayon moisture-resisting membrane placed between the brass plate and the diaphragm.

The shape of the electrodes and the carbon chamber provides sufficient contact force between the diaphragm electrode and the granular carbon in the zone of maximum current density so that this transmitter operates satisfactorily in any position. When new, it has a resistance of around 30 to 40 ohms.

9.5 Resistors, Inductors and Capacitors

A resistor is a piece of apparatus designed for the specific purpose of inserting resistance into an electric circuit. Resistors are therefore designed in practice to have a definite value of resistance in ohms. They must also be designed in many cases to be capable of dissipating specified amounts of power without heating beyond safe limits. The applications of resistors in telephone and telegraph work are almost innumerable, involving ohmic values ranging from fractions of one ohm to many megohms; and power ratings (i.e., ability
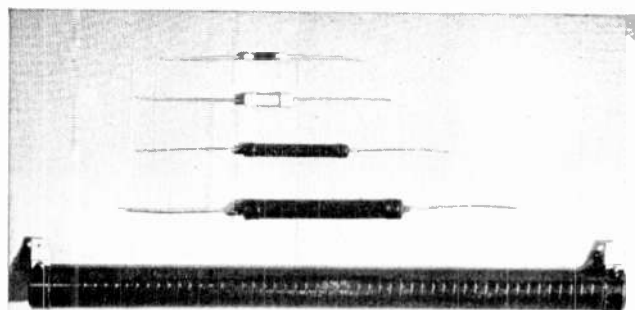
DEPOSITED CARBON RESISTORS

to dissipate energy) ranging in general up to about 5 watts.

The great majority of the resistors used can be divided into two principal types—wire-wound and carbon. For relatively low resistances ranging from less than one ohm up to a few thousand ohms, most of the common resistors found in the telephone plant are the so-called flat types coded #18 and #19 and illustrated in Figure 9-12. As indicated, these consist of wire wound on a flat card of insulating material, the flat shape having the advantage of making the resistor relatively non-inductive. Where higher resistances—in the order of ten thousand ohms and upward—are required, some type of carbon resistor is usually employed. Most common of these is the so-called **composition** resistor in which the resistive element is a combination of finely divided carbon or graphite mixed with a non-conducting filler such as talc, with synthetic resin as a binder. These resistors are well suited for application in relatively high frequency circuits such as radio where great precision in ohmic value ordinarily is not required. They are usually built to tolerances of about 5% and are commercially available in values from 10 ohms to 22 megohms, with power ratings up to 2 watts. Where very high resistances are needed, the **deposited-carbon** resistor may be used. In this type the resistor element is an extremely thin film of carbon, or a mixture of carbon and boron, deposited on the surface of a ceramic core by thermodecomposition. A helical groove is then cut through the carbon film to leave a ribbon of the carbon film wound around the core between the terminal electrodes at the ends.

The resistance of any ordinary electric conductor, such as the wire used in wire-wound resistors increases as the temperature of the conductor increases. There are, however, many so-called semiconducting materials in which the opposite effect

18 - Type   19 - Type
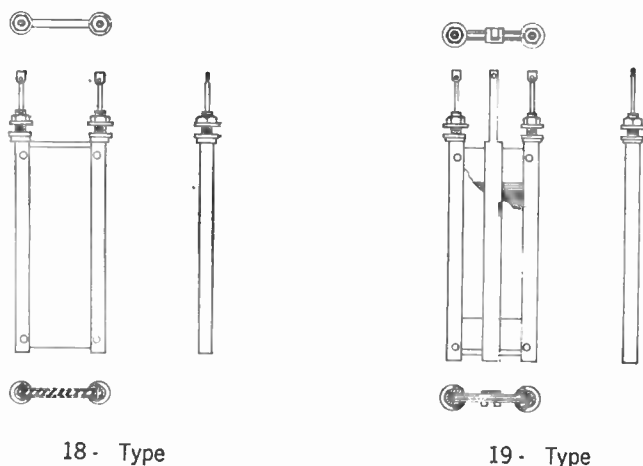
FIG. 9-12 FLAT RESISTORS

[69]

occurs—that is, the resistance decreases as the temperature increases. Such materials include the oxides of manganese, nickel, cobalt, iron, and zinc, or mixtures of such oxides. Resistors made up of these materials are known as **thermistors**. They have many important applications in telephone work, particularly in controlling amplifier gains to compensate automatically for changes in line loss resulting from temperature changes along the line.

An inductor is a piece of apparatus designed primarily to insert inductance into a circuit. We have already noted the use of certain types of inductors, commonly called retardation or choke coils, in power circuits. Inductors, as was pointed out in an earlier Chapter, consist fundamentally of coils of wire, the amount of their inductance depending primarily upon the number of turns in the coil, its size and shape, and the nature of the core material about which the coil is wound. Inductors are usually classified under two main types —those having air cores and those with cores consisting of iron or iron alloys. Both types have numerous uses in communication circuits. As would be expected from our knowledge of magnetism, air core inductors have relatively low inductance values but are comparatively free from the core losses due to magnetic hysteresis and "eddy currents" that are inevitable when metallic cores are used. The principal field of use of air core inductors is in high frequency or radio work. Where relatively large amounts of inductance are needed, as in power circuits and telephone and telegraph circuits at ordinary frequencies, iron core inductors are employed.

Since any inductor consists of many turns of wire, it will always have resistance as well as inductance. In inductor design, however, every effort is made to keep the resistive effect as low as possible in comparison to the inductive effect. The ratio of the inductive effect to the resistive effect is known as the **figure of merit** or "Q" of the inductor.

Capacitors (or condensers) were discussed briefly in Chapter 8. It was established then, that the capacitance value depends on the plate areas, and the nature and thickness of the dielectric between the plates. The simplest dielectric is air, and most of the variable capacitors are "air condensers". Unless such capacitors are extremely large, however, their total capacitance is relatively low and their chief use is, accordingly, in high frequency work. For the many applications in communications work at lower frequencies, capacitors employing impregnated paper, mica, or ceramic materials as dielectrics are extensively used. The paper capacitor may be made as a continuous roll of two aluminum foil strips separated by thin Kraft paper impregnated with oil or wax. Or, to reduce its volume, metal films may be evaporated directly onto the paper strips. Such capacitors are made in a wide range of capacitance values extending up to 5 microfarads and may be designed to withstand substantial voltages (up to 4000 volts) without breakdown.

Mica capacitors are made in a smaller range of capacitance values. This is also true of the ceramic capacitors, which employ materials based on titanium dioxide mixtures as dielectrics. Neither will withstand as high a voltage as good paper capacitors but both have a higher degree of stability with respect to temperature change and aging. Their normal application, accordingly, is to situations where more precise and constant capacitance values are required.

Within reasonable limits of cost and size, none of the capacitors discussed above can be built to have capacitance values exceeding a very few microfarads. Where larger values of capacitance are needed, an entirely different type known as an electrolytic capacitor is widely used. In most designs this capacitor consists of plates of aluminum or tantalum, on which extremely thin oxide films have been electrochemically deposited, immersed in a conducting aqueous electrolyte. The aluminum or tantalum plate is one electrode of the capacitor, the conducting electrolyte is the other electrode, and the insulating oxide film on the plate is the dielectric. Because of the extreme thinness of this film, capacitors of this type may have a very high capacitance even though quite
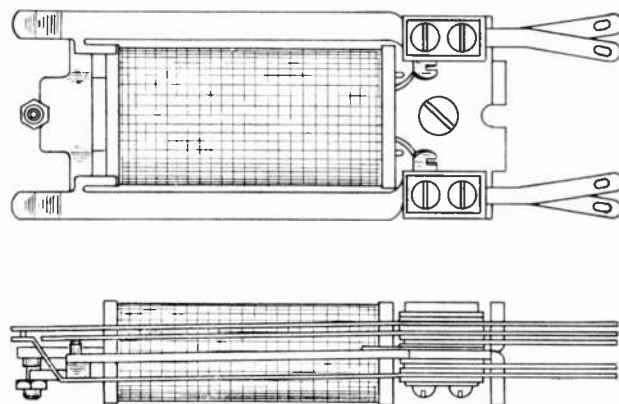
FIG. 9-13 E-TYPE RELAY
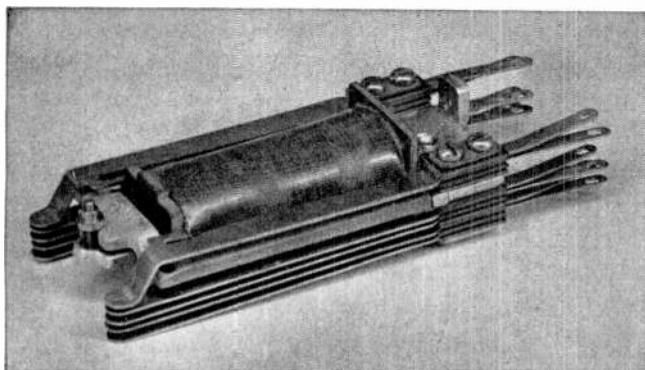
E-TYPE RELAY

limited in size. For certain applications in telephone power plants, such as in voltage stabilizing filters, electrolytic capacitors may be built in large cells similar in external appearance to a storage battery. For much more extensive use in electronic devices such as amplifiers, they are usually built in small cylindrical forms not dissimilar in appearance to the small paper capacitors commonly employed in such apparatus. The electrolyte in this case is usually a viscous paste which includes some glycol or other conducting liquid. These capacitors are generally not as stable or long-lasting as paper or mica capacitors, principally because the partly liquid electrolyte offers a possibility for undesired chemical action if any impurities are present. There is a design of tantalum capacitor, however, in which the "electrolyte" is dry, solid manganese dioxide, with the metal container acting as one electrode. Since it is therefore chemically inert, this design makes for longer life and greater stability.

9.6 Relays and Switches

The fundamental instruments of telephony are the telephone transmitters and receivers that are described in some detail in a preceding Article. These instruments, when associated in appropriate circuits with other basic apparatus units such as coils, capacitors, resistors, etc. and connected together by means of appropriate transmission lines, make it possible for persons at different locations to talk with each other.

The problem of telephone service, however, is not just to make it possible for some one person to talk with another person at a distance, but for any subscriber to such a service to be able to talk with any other subscriber. This means that any practical working telephone plant must include vast quantities of switching apparatus designed to connect, disconnect and rearrange a great variety of circuits quickly and surely, as may be necessary to meet the ever-changing communication requirements of the subscribers.

Switching apparatus may be classified first into two fundamental categories—manually operated and electrically operated. Devices in the first category include jacks, plugs, cords, keys, push buttons, etc. Such devices are probably too well known to require any extended discussion here, although the reader who is familiar with them will also recognize that they can be, and are, built in many different designs to serve particular purposes.

The most commonly used of the electrically operated switching devices is the relay. This is a
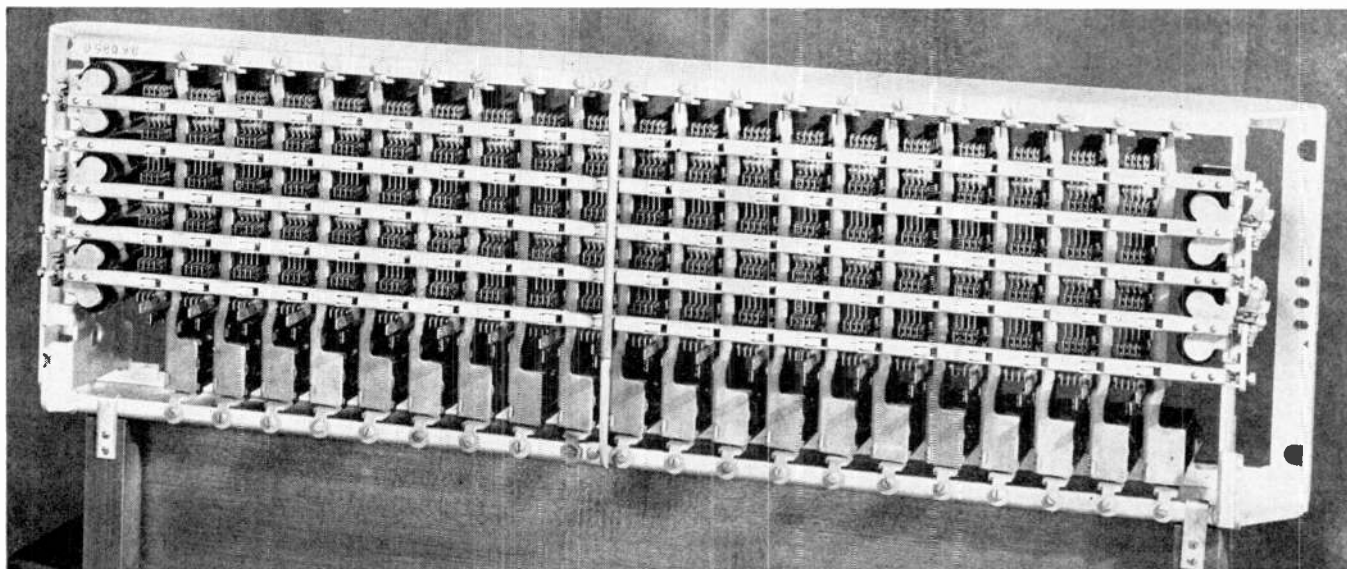
200-POINT CROSSBAR SWITCH

relatively simple piece of apparatus in which an electromagnet driven by one electric circuit may operate to open or close switching contacts in one or more other electric circuits. Figure 9-13 and the accompanying photograph illustrate the basic design of a general type of relay in wide use in the Bell System. Here pairs of electrical contacts, normally held closed by the tension of flat springs, may be opened by the movement of the armature that results from energizing the magnet coil winding. With other arrangements of the springs, normally opened contacts may be closed by operation of the relay armature; different sets of contacts may be opened and closed simultaneously, or in a desired order. Indeed, since relays of this type are built with as many as 24 springs, the number of possible circuit arrangements is very large and runs well up into the hundreds in standard telephone practice.

The relay shown in Figure 9-13 is equipped, as previously noted, with contact springs made of flat metal strips. Most of the relays currently manufactured by the Western Electric Company employ a pair of wire springs in place of the older flat spring, as shown in the accompanying photograph. To each wire is welded a contact point, usually of some precious metal to insure a good electrical connection. The twin contacts guard further against the possibility of faulty operation in the event of dust or other foreign particles lodging between one pair of contact points. Wire spring relays are built with the same variety of open and closed contact arrangements as the older types. Their construction, however, lends itself to better quality control in manufacture and to substantial cost savings.
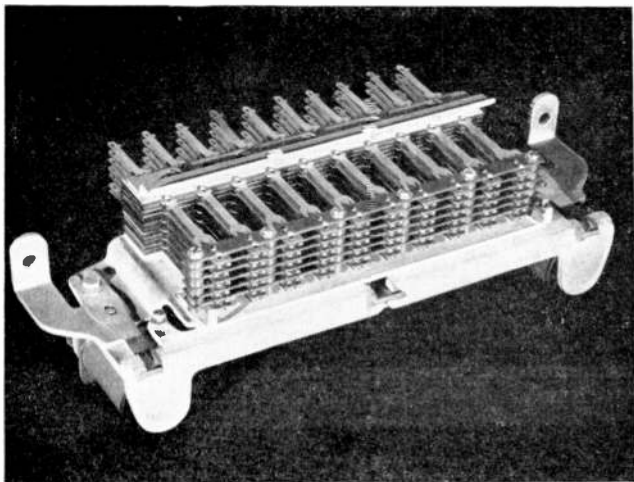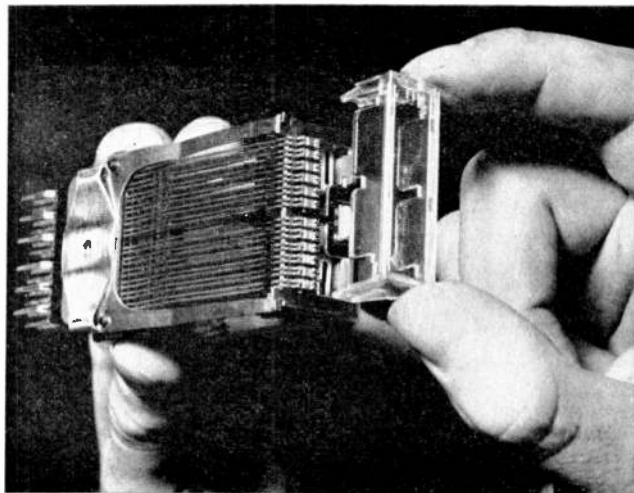
WIRE-SPRING RELAY

Since relays are mechanical devices having moving parts with inertia and elasticity, their operation cannot be instantaneous. The operating time of the ordinary relay is in the order of 50 milliseconds, either on opening or closing. In some telephone circuits, however, it is desirable to control relay operating time in such a way that it will be longer than the nominal time. This can be effected within limits by adding a very low resistance path around the magnetic core of the relay in the form of either a short-circuited winding or a solid metal sleeve. Then, when the current in the regular relay winding is broken, the decaying flux in the core will induce a circulating current in the sleeve. This, in turn, tends to produce flux in the core additional to and in the same direction as the original flux. The time required for the total flux to decay is thus increased, which causes a proportionate delay in the relay release time. The lower the resistance of the added sleeve or short-circuited winding, the greater will be the current and the greater the retarding effect on the relay release time. Thus by varying the resistance of the sleeve, or the short-circuited winding, it is possible to obtain a range of delay times extending from a minimum of 50 ms up to as much as 500 milliseconds.

Where, as may frequently be the case in the operation of local or toll dialing machinery, it is necessary to operate more than a dozen or so switching contacts simultaneously and under one control, a special type of relay known as a **multi-contact** relay is used. As shown in the photograph, this relay has two coils and two armatures each of which operates half the contacts. The springs are arranged in ten groups, each of which

MULTI-CONTACT RELAY

may include as many as six pairs of springs. This provides for a maximum of sixty simultaneous contacts or switching operations when the two armatures are operated together. The complete structure may be used as two independent relays, each having up to thirty contact springs, or as a single relay with double this capacity when the operating coils are connected in parallel. All contacts are doubled to provide greater reliability of operation.

All relays used in telephone circuits are designed to operate and release at certain definite values of current in their windings. As these current values are frequently very small, this means that the springs which hold the armatures in their non-operated positions must be adjusted with precision. Practically all telephone circuits are
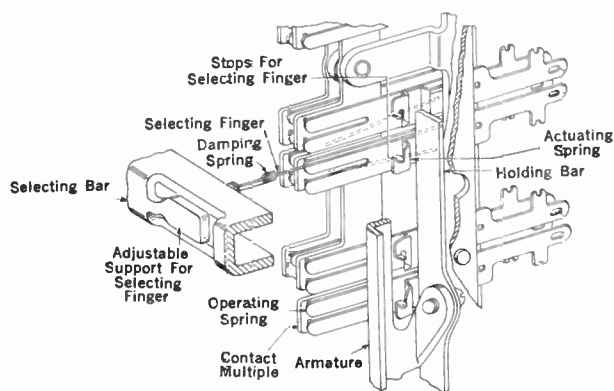
FIG. 9-14 CROSSBAR SWITCH CROSSPOINT

dependent for their operation upon the proper functioning of relays. In some of these the operating limits are sufficiently liberal to allow considerable margin in adjustments. But in others—and these are frequently the more important ones—the difference between an adjustment giving satisfactory operation and one under which the relay will fail to function properly, may be very small. In practice, specific instructions giving the exact operating and release current values for which each type of relay should be adjusted for each kind of circuit in which it may be used, are provided.

In many of the more complicated switching operations required in telephone and telegraph work, such as those involved in toll line dialing for example, it is necessary to operate switches

which are electrically controlled by two rather than by only one external circuit as in the case of the relays already discussed. A commonly used switch of this type, which has many of the physical characteristics of a relay, is known as the **crossbar switch.** Such a switch arranged to provide a maximum of 200 different switching connections is shown in the accompanying photograph. (Note, however, that not all of these connections can be made simultaneously, but at each of the switch-points as many as six different contacts may be operated.) The 200 switching points or **crosspoints** of the switch shown are obtained by the use of 20 vertical bars and 5 horizontal crossbars.

Closure of any one crosspoint is effected by the coordinated operation of one of the horizontal and one of the vertical bars. As shown in the photograph, the horizontal and vertical bars are located in front of the contacts and arranged to be rotated by magnets through a small arc. Each of the vertical bars is provided with a magnet at the bottom of the switch. Only five horizontal bars are used but each bar is equipped with two magnets so that it may be rotated in two directions.

Figure 9-14 shows in more detail how the controlling magnets operate the crosspoint switches. Each horizontal or **selecting** bar carries twenty selecting fingers which are normally opposite an open space between two crosspoints. When one of the bars is partially rotated by its selecting magnet so that the fingers are tilted upward or downward, its twenty fingers lie across the backs of the contacts as indicated by the dotted lines in the drawing. If one of the vertical bars is now rotated so as to press against the vertical row of five selecting fingers, the selecting fingers that are in normal position will be pressed into the spaces between crosspoints, with no effect. The selecting finger that has been tilted upward or downward by the operation of the horizontal bar, however, will be pressed against the adjoining flexible contact spring assembly, thus closing the contacts at that crosspoint. Once this contact has been made in this way, it will remain closed as long as the vertical bar is held in the operated position even though the horizontal bar has been returned to normal, because the selecting finger will remain trapped against the contact spring. For this reason the vertical bar is known as **a holding** bar and its associated magnet is known as a **holding magnet.**

[73]

CHAPTER 10

TELEPHONE CIRCUITS

10.1 The Telephone Subscriber Set

Figure 9-10 of the preceding Chapter showed a simple telephone connection between two telephone sets, each equipped with a transmitter, receiver, induction coil and its own battery for supplying talking power. In most modern telephone station installations, talking battery is supplied to each subset from a common battery at the telephone central office to which each subscriber line is connected. The simplest subscriber station circuit arrangement under these conditions is shown schematically in Figure 10-1. When the receiver is lifted to close the contacts of the hook-switch, and the line is picked up at the central office by an operator or mechanical device, the central office battery is connected in series with the primary winding of the induction coil and the transmitter, and current is sent over the line.
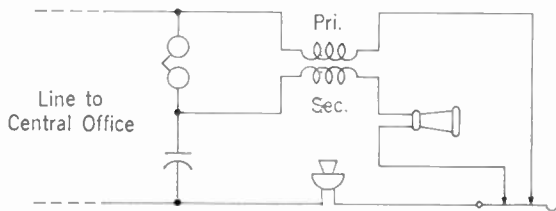
FIG. 10-1 SIMPLE STATION CIRCUIT

Varying currents set up by the transmitter, when it is talked into, add to or modulate the direct current flowing from the central office battery. There is also a path for the varying transmitter currents (which are fundamentally alternating rather than direct) through the capacitor, the secondary of the induction coil and the receiver. Thus, there are varying currents in both windings of the induction coil and each induces a voltage, and a consequent current, in the other winding. The two currents in the primary winding combine to flow out on the line and thence to the distant receiver, and the two currents in the secondary winding combine and flow into the receiver and cause sidetone.

When receiving, the incoming varying (alternating) current, which carries the message, di-vides between the two parallel paths containing the transmitter and receiver respectively. In passing through the windings of the induction coil, in these two paths, the current in the primary winding induces a voltage in the secondary winding, which opposes the part of the incoming current flowing into that winding; but the circuit is so designed that the part of the incoming current in the secondary winding is greater than that in the primary so that the resultant current flowing in the receiver is still sufficient to operate it. Since in receiving, a substantial part of the incoming energy is dissipated in the transmitter circuit and in transmitting, energy is likewise dissipated to no useful purpose in the receiver circuit, the subset is fundamentally inefficient. Such inefficiency, however, is inherent in any circuit designed to operate in two directions without using directional switching arrangements.

The transmitting current flowing in the receiver as sidetone obviously serves no useful purpose and may be confusing or annoying to the speaker. The simple subset discussed above is modified generally in modern practice to a circuit arrangement such as that shown in Figure 10-2, which is one of several possible "anti-sidetone" connections. This circuit makes use of a different induction coil having a third winding S_2, the resistance of which is made high relative to that of winding S_1. For the receiving condition the current relationships are effectively identical with those of the sidetone circuit shown in Figure 10-1 because the resistance of S_2 is so high. For the transmitting condition there are, as in the sidetone circuit, two parallel paths for the current
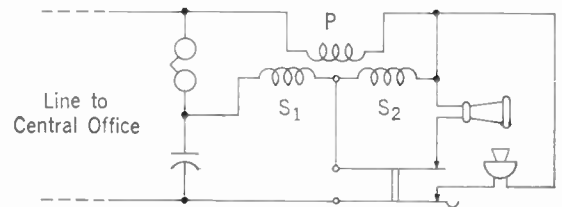
FIG. 10-2 ANTI-SIDETONE STATION CIRCUIT

flow—one directly over the line through the primary of the induction coil and the other through the capacitor and S_1 and either through S_2 or the receiver. Because of the relatively high resistance of S_2 most of the current will tend to flow through the receiver, but the induction coil is so designed that current flowing in P and S_1 will induce a voltage in the winding S_2 which will tend to set up a current approximately equal and opposite in direction to that flowing through the receiver. The net effect is no current in the receiver, and hence no sidetone. This circuit is no more efficient than the sidetone circuit but it seems more so, principally because in receiving the listener is not distracted by extraneous noises coming from his own transmitter.

10.2 The Telephone Central Office

In practical telephone systems all subscriber stations are connected to a telephone "central office" by relatively short lines known as subscriber loops. The major purpose of a central office is to provide switching arrangements whereby any subscriber station connected to that office can be connected at will to lines leading to other subscriber stations also connected to that office; or to trunk lines leading to other central offices in the same or distant cities. The central office has numerous incidental functions, including that of providing common talking battery for all of its lines as discussed in the preceding Article. This requires comparatively large installations of storage batteries, power generating machinery and associated control equipment. Installed in the central office too, are numerous frames and racks, testboard and other testing equipment, signaling
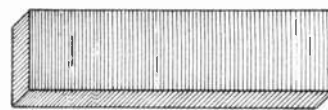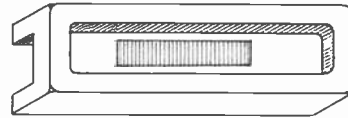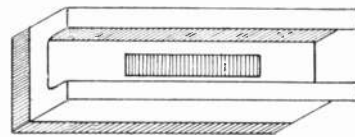
FIG. 10-3 OPEN-SPACE CUTOUT

and control devices, and various other auxiliary equipment. Important, both at the central office and the subscribers' stations, is the installation of protective devices which will guard both telephone apparatus and people using it from dangerous foreign voltages that may be brought in on the lines as the result of some abnormal condition. This protection is of such basic importance that it may be well to digress here to review it in some detail.

Practically all outside telephone plant, except conductors that are completely underground from terminal to terminal, may be occasionally exposed to excessive voltages from such sources as lightning and other atmospheric disturbances, electric power lines running in close proximity to the telephone lines, high power radio sending apparatus, etc. Accordingly, such exposed conductors leading into a central office or subscriber station are connected first to appropriate protective apparatus. The particular protective units employed and the manner in which they are connected into the telephone circuits vary somewhat with particular situations, but in general protective devices are of three principal types—open-space cutouts, fuses, and heat coils.

The first and last of these devices ordinarily operate to ground the protected wire, while the fuse opens the wire in which it is inserted. Each of the protective units is designed so that, for the particular situation in which it is used, it will be

500-TYPE TELEPHONE SET
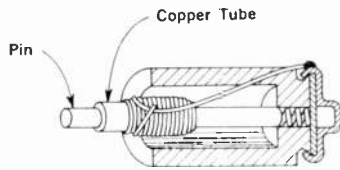
Fig. 10-4 Heat Coil.

sufficiently sensitive to operate before the plant which it is protecting is damaged, but on the other hand, not so sensitive as to cause an unnecessary number of service interruptions.

The standard form of open-space cutout used at subscribers' stations, in central offices, and at the junctions of cable and open wire lines, is illustrated in Figure 10-3. It consists of two carbon blocks having an accurately gaged separation of a few thousandths of an inch, one of which is connected to ground and the other to the wire to be protected. As shown in the Figure, one of the carbon blocks is much smaller than the other and is mounted in the center of a porcelain block. When the voltage of the telephone wire becomes too high, the wire will be grounded by arcing across the small air-gap between the carbon blocks. If a considerable current flows across the gap in this way, enough carbon may be pulled from the blocks by the arc to partially fill in the gap and cause permament grounding. Or, in the extreme case, when the discharge is prolonged and sufficiently high, the glass cement with which the small carbon insert is held in the porcelain block may be melted, with the result that the blocks are forced into direct contact by the mounting springs in which they are held. In the majority of protector operations, however, the blocks do not become permanently grounded.

The air-gap space between the blocks is designed so that the operating voltage of the protector will be less than the breakdown voltage of the weakest point of the circuit which it is designed to protect, and greater than the maximum working voltage of the circuit. The average operating voltage of the open-space cutouts used at subscribers' stations and in central offices is about 350 volts. For the cutouts used at junctions between open wire and cable lines, an average operating voltage of about 710 volts is standard.

When a telephone conductor is grounded by the operation of an open-space cutout, current will continue to flow through the telephone conductor to ground so long as the exposure continues. This current may be large enough to damage the tele-

phone conductor or the protective apparatus itself. Accordingly, it is necessary to insert in the conductor, on the line side of the open-space cutout, a device which will open the conductor when the current is too large. Fuses are used for this purpose. The fuse is simply a metal conductor inserted in series with the wire to be protected, which is made of an alloy or of a very fine copper wire that will melt at a comparatively low temperature. Short lengths of cable conductors (six feet or more) of 24 or finer gage will serve effectively as fuses and will fuse on current values not high enough to overheat dangerously the central office protectors. Where the use of such inserted fine gage cable is not practicable, lead alloy fuses mounted in fire-proof containers or on fire-proof panels are employed. These are also designed to operate with a current of 7 to 10 amperes.

Finally, it is frequently necessary to protect telephone apparatus against external effects in which the voltage is not high enough to operate the open-space cutout, nor the current high enough to operate fuses, but still high enough to damage apparatus if allowed to flow over a long period. Such currents are usually called "sneak" currents and are guarded against by the use of heat coils. As illustrated in Figure 10-4, the heat coil consists of a small coil of wire wound around a cop-
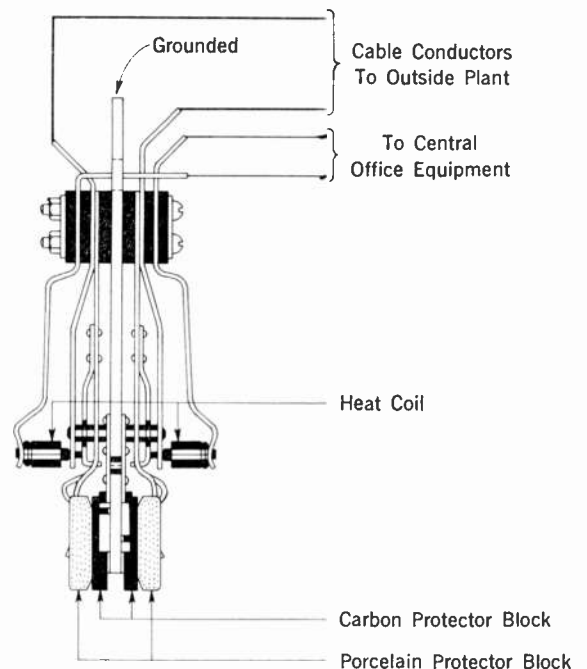
Fig. 10-5 Heat Coils and Cutouts Mounted on Protector Frame in Central Office

[76]
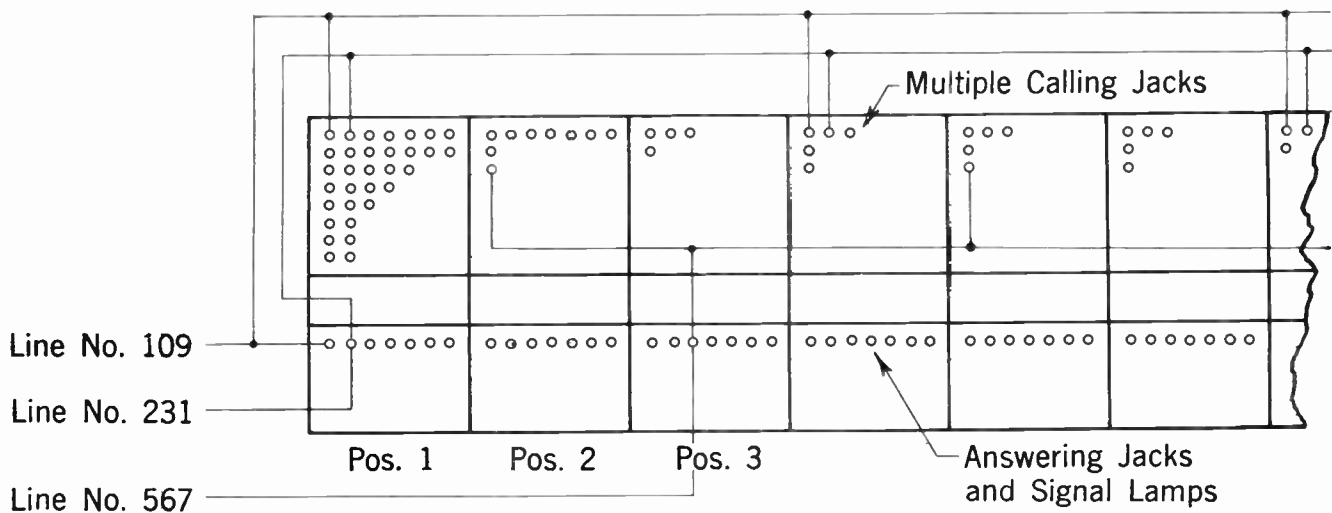
Fig. 10-6 Multiple Switchboard

Line No. 109

Line No. 231

Line No. 567

Pos. 1   Pos. 2   Pos. 3

Multiple Calling Jacks

Answering Jacks
and Signal Lamps

per tube which is connected in series with the wire to be protected. Inserted within the copper tube and held in place by an easily melting solder is a metal pin which is connected to the line side of the coil. If sufficient current flows through the coil to melt the solder, this pin will move under the pressure of its mounting spring and thus connect the line to ground. The heat coil now in general use in the telephone plant is designed to carry .35 ampere for three hours and to operate in 210 seconds on a current of .54 ampere. In certain cases heat coils of a generally similar nature are used to open circuits instead of to ground them. Where used in line circuits, as in the case of conductors entering a central office, the heat coil is mounted on the office side of the open-space cutout. In this position the heat coil wiring aids the operation of the open-space cutout by presenting a considerable resistance to suddenly applied voltages such as are produced by lightning discharges. The method of mounting heat coils and open-space cutouts on the protector frames in central offices is illustrated in Figure 10-5.

Let us return now to the switching function of the central office, which we have already pointed out as its main purpose. Switching equipment is designated as **manual** or **dial** depending upon whether the switching operations are performed primarily by people or by mechanical devices. Although both types perform the same function, the equipment and methods employed in each case are quite different.

10.3 Manual Switching Systems

In the manual central office, switching is performed by operators at switchboards. In a local single office exchange, all subscriber lines terminate in at least two jacks in these switchboards. One jack, mounted in the lower part of the switchboard panel, has associated with it a signal lamp by means of which a subscriber wishing to make a call can secure the attention of the operator. This is her **answering jack.** The other jack mounted in the upper panel of the board, is used by the operator to complete a connection to this same subscriber's line when it is called by some other station. This is designated the **calling jack.** Except in very small offices, each line appears in parallel at many other jack locations along the board. In other words, the terminations are **multipled** so that any line is within reach of any operator.

The principle of the multiple switchboard is that the answering jacks and associated signals are divided up among the various operators, each operator handling on the average about two hundred lines and being responsible for answering any signals from these subscribers. In addition to these answering jacks, there may be as many as 3,300 calling jacks in the position in front of each operator. These calling jacks do not have any signals mounted with them, as they are for calling only. The calling jacks are each multipled, that is, connected in parallel, with a similarly located jack in the third position to the left and right. Any operator can reach any one of about

[77]

10,000 calling jacks, either directly in front of her or in the adjacent positions on her left or right. A multiple switchboard is shown diagrammatically in Figure 10-6. In this Figure should subscriber Number 109 call subscriber Number 567, the signal would come in at position "1" where the answering jack for subscriber 109 is located and the operator would connect him by plugging into calling jack Number 567 in the multiple to her right (Position 2). On the other hand, if subscriber 567 called subscriber 109, the operator at position 3 would answer his call and connect him to subscriber 109 by means of the calling jack in the multiple to her right (Position 4). Each operator is warned against plugging into a busy line by means of a "click" which is heard in her head receiver when she starts to plug into a calling jack already in use somewhere else in the multiple.

Figure 10-7 shows a telephone connection between two stations terminating at the same central office. Here the telephone circuit at each station is normally open when the receiver is on the hook, with the exception of the ringer which is bridged across the circuit in series with a capacitor. It is a function of the capacitor to close the circuit for alternating current and open it for direct current. Accordingly, the line is open in so far as the subscriber's signaling the operator is concerned and is closed through the ringer in so far as the operator's ringing the subscriber is concerned; or we may say, the circuit is in such condition that the subscriber may call the operator or the operator may call the subscriber at will. The subscriber calls the operator by merely closing the line, which is accomplished by removing the receiver from the hook. Contacts C_1 and C_2 are made at the hook switch. C_1 closes the line through the transmitter in series with the primary of the induction coil. This permits current to flow from the central office battery B through one-half of the line relay winding R_1, over one side of the line L_1, through the primary winding of the induction coil, and the transmitter back to the central office over the other half of the line L_2, through the other half of the relay winding R_1, to ground. This energizes the line relay R_1 which connects the central office battery to the answering lamp A_1 in the face of the switchboard in front of one or more operators. This lamp, lighting, indicates that this particular line is calling. The operator answers the call by inserting plug P_1 into the answering jack associated with the

lighted lamp and to which the line of the calling party is connected. A third battery connection to the sleeve of the plug closes a circuit through the winding of a second relay R_2, known as a "cut-off" relay, which disconnects the line relay from the circuit, putting out the burning answering (or line) lamp A_1. The operator learns the calling subscriber's wishes by connecting her telephone set to the cord circuit by means of the key K_1. She talks over the two heavy conductors of the cord circuit through the windings of the repeating coil, which by means of transformer action induces current into the other windings of the same coil; this flows back over the calling subscriber's line and induces a current in the secondary of the induction coil, which flows through the telephone receiver.

Not only does the operator's voice current flow from the central office cord circuit to the subscriber's receiver, but there is direct current furnished by the central office battery through two of the four windings of the repeating coil of the cord circuit, over the line, and through the subscriber's transmitter. This permits the subscriber to talk by virtue of the transmitter carbon resistance varying the strength of the current, which, by means of the repeating coil windings at the central office, induces an alternating voice current across to the opposite side of the cord circuit.

Upon learning the number of the party called, the operator inserts plug P_2 into calling jack J_2 which permits the lamp S_2 to burn because the circuit is closed from the central office battery through the sleeve connection and the cut-off relay winding. This lamp tells her that the receiver is on the hook at the called party's station and that she must give this connection attention by ringing the called party at frequent intervals. This is accomplished by operating the ringing key K_2. When the called party answers, current flowing from the central office battery through the windings of the repeating coil, and through the supervisory relay R_3, operates this relay. As a result the lamp S_2 is short-circuited and goes out, notifying the operator that the party has answered. At the same time, a resistance is inserted in the battery circuit to limit the current through the cut-off relay. When both parties finish talking and hang up their receivers, this supervisory relay, as well as the corresponding relay on the other side of the cord circuit, is de-energized, and since the short-circuit is then removed from the lamps,
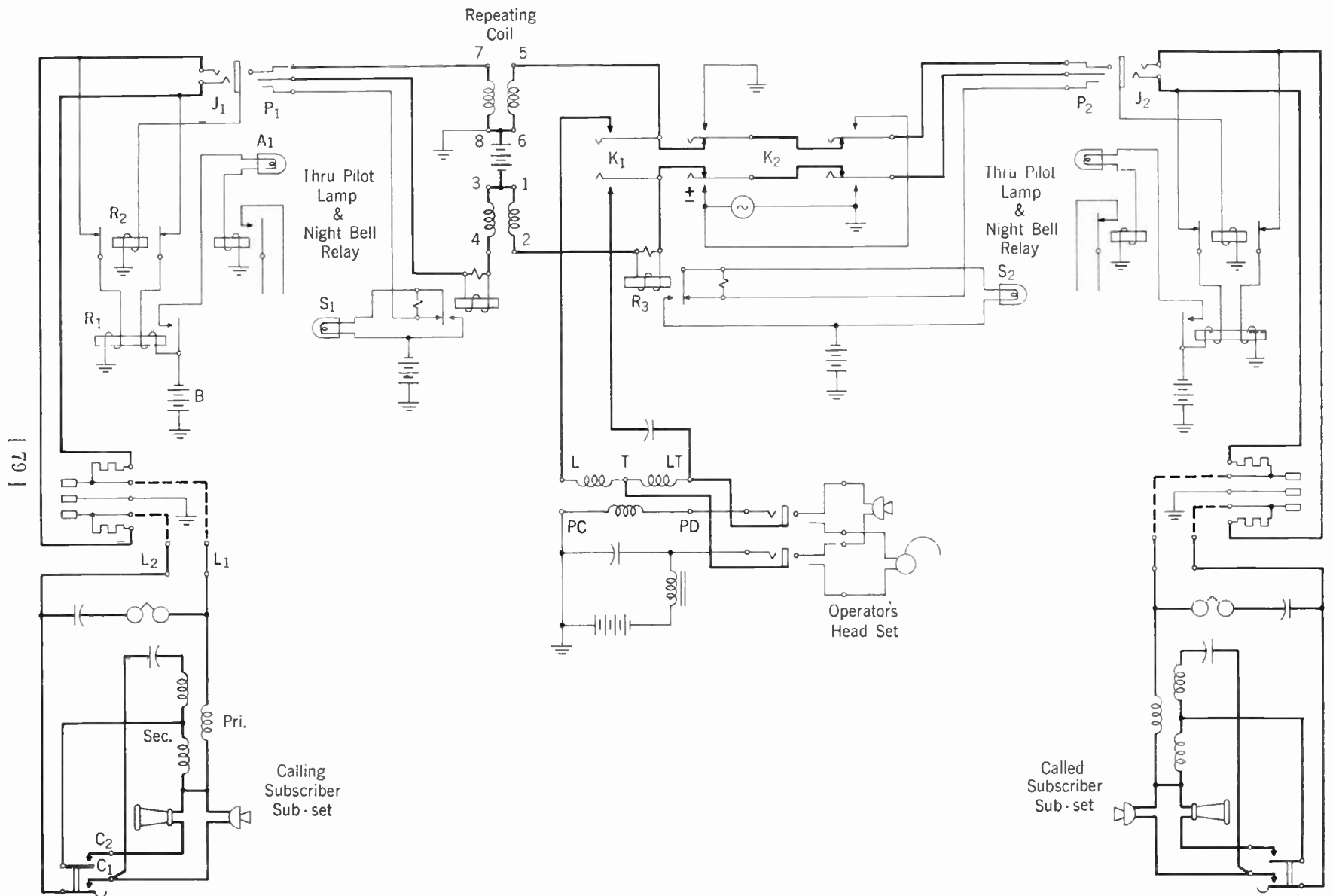
Fig. 10-7 Telephone Connection Through Common Battery Exchange
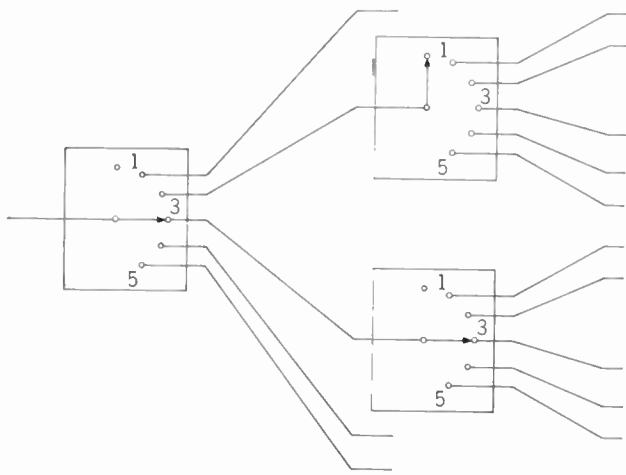
FIG. 10-8 PRINCIPLE OF STEP-BY-STEP SWITCHING

they light. This notifies the operator that both parties are through talking and that both cords are to be taken down. When the operator pulls down both cords, the sleeve circuits of the cords are opened at the jacks and the lamps go out.

What we have discussed above is perhaps the simplest example of a telephone switching operation—that is, a connection between two subscribers of the same single exchange established by one operator at an "A" switchboard. Where a city or other exchange area is of such size as to require more than one central office, a connection between a subscriber whose line terminates in one office and a subscriber whose line terminates in another office cannot be completed by one operator. In such case the "A" operator who picks up the call from a calling subscriber, uses her cord circuit to connect the calling subscriber's line to a trunk leading to a switchboard in the other central office in which the called subscriber's line is terminated. This switchboard is known as a "B" board. It differs somewhat in appearance and in the detail of its circuit arrangements from the "A" board. The incoming trunks are terminated in the "B" operators' key shelf in plugs, and the panels of the switchboard are used almost entirely for calling or multiple jacks.

When the "A" operator at a distant office connects a calling line to a trunk to this "B" board, a lamp associated with the plug and cord in which the trunk terminates in the "B" board key-shelf will light. At the same time the "B" operator's head-set is automatically connected so that the "A" operator at the calling office can pass her the number. The "B" operator then inserts the plug into the called jack. Ringing is usually automatic

and the supervisory signal lamp at the "B" board remains lighted until the called subscriber answers. After the conversation is completed and the called subscriber hangs up, the supervisory lamp again lights indicating to the "B" operator that she should take down the connection.

There are other situations in large cities where it is uneconomical to provide direct switching trunks from each central office to every other central office in the exchange area. In these cases a third type of switchboard known as a "tandem" board is provided in some appropriately located intermediate central office. A connection then requires the service of a third operator. The "A" operator at a calling office connects to a trunk leading to the tandem board at the intermediate office; the operator there connects this trunk to another trunk leading to the called office; and the "B" operator there in turn connects the trunk to the called subscriber's line. Naturally, the switching circuits in these three types of switchboards vary in detail, particularly in respect to the supervisory signaling arrangements. Basic methods
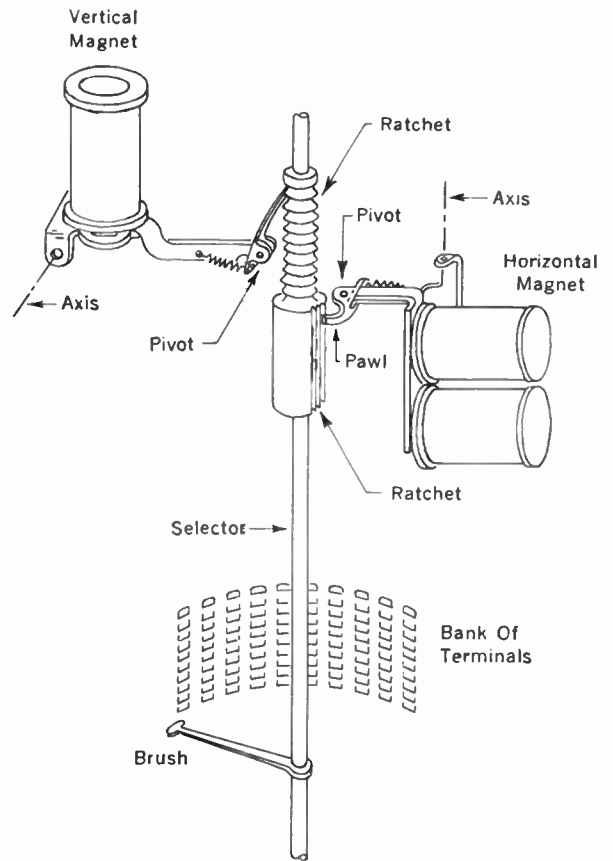
FIG. 10-9 STEP-BY-STEP LINE-FINDER SWITCH

and principles, however, do not differ fundamentally from those described in connection with Figure 10-7.

The requirements for long distance or toll switching are essentially similar to those for local switching. Again, there are three principal switching situations—between the calling subscriber and the toll trunk or line, between the toll trunk and the called subscriber, and where necessary, between two toll trunks. In long distance practice, these operations are performed at outward, inward and through switchboards, respectively. Signaling and supervisory facilities are necessarily somewhat different in toll switching than in local switching. In general, the toll switchboard must be designed to permit a larger variety of switching operations and the cord circuits and other equipment facilities associated with it are accordingly somewhat more extensive. Because of the variety of switching situations, there is no one specific design for toll switchboard arrangements that applies alike to all such switchboards. The reader who wishes to study switchboard circuit arrangements in detail may refer to the standard drawings and instructions applying to the particular switchboards in which he is interested.

10.4 Dial Switching Systems

The fundamental operations to be performed by a mechanical switching system are naturally the same as those discussed above. It is hardly necessary to point out, therefore, that any completely mechanical system must be made up of a considerable quantity of intricate electro-mechanical equipment. In the local dial exchange the operation of the mechanical switching system is controlled by a dial installed on the subscriber's telephone set, which transmits to the central office electrical impulses corresponding to each digit dialed by the subscriber.

The oldest type of dial central office machinery in use functions on a "step-by-step" basis. This means that each digit dialed causes the movement of a switch at the central office which will carry the connection a step forward toward its ultimate objective—the called subscriber's line. The basic principle involved can be readily understood by examination of the simplified example shown in Figure 10-8. Here only two steps are shown and there are only five contacts on the switches of each step. If the switches are arranged to move in order through the same number of contacts as the number of electric impulses

coming from the dial, any one of twenty-five separate telephones could be reached by dialing only two successive digits between 1 and 5. If, as in the actual telephone dial, there are 10 digits and the system is designed for dialing as many as seven digits to reach the desired telephone, the simple diagram of Figure 10-8 could be expanded without any change in principle to show how any one of ten million telephones could be reached in the seven successive dialing steps. The trouble with such a simple arrangement as this is that it would require a tremendous number of switches and connecting wires, and still permit only one telephone to make a call to any of the others.

By suitable arrangements, however, this principle can be employed to permit the origination of the call by any of the telephones and still keep the total amount of switches and other equipment within reasonable bounds. The first step is to eliminate the necessity for having a separate selecting switch associated with each subscriber's line. Instead, switches known as **line-finders** are provided for groups of subscriber lines, each one of which is connected in multiple to the "bank" terminals of the switches. In the **step-by-step** system, the line-finder switch resembles that illustrated in Figure 10-9, which is arranged to move vertically through ten steps and horizontally through ten steps for each vertical step. When a subscriber picks up his receiver, a relay associated with his line causes an idle line-finder switch to go into operation and hunt for the terminal to which his line is connected.

With this arrangement, a simple dial system might take the form illustrated in Figure 10-10. Here the line-finder switch connects the calling subscriber's line to a trunk selecting switch which is installed in the same central office, and in the step-by-step system is of the same design as the line-finder switch illustrated in Figure 10-9. Dial tone is automatically sent back to the calling subscriber and as he begins to dial, the trunk selector switch in office "A" moves to a terminal at which a trunk corresponding to the initial digit dialed is terminated. This trunk may lead to a distant office or to terminating equipment in the same office. In the simple case illustrated, the first switch at office "B" selects a group of connecting lines under the direction of the second number dialed by the calling subscriber and moves the call on to a final selecting switch, which is operated by the third number dialed to connect to the called subscriber's line.
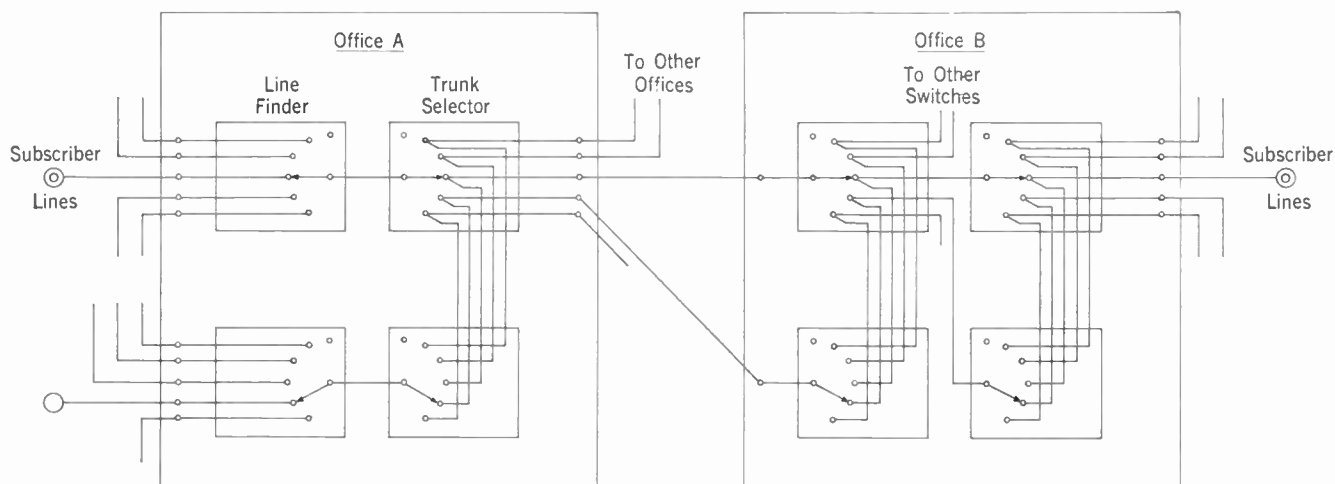
FIG. 10-10 INTER-OFFICE DIAL SWITCHING PRINCIPLE

The above illustration of course is over-simplified as compared to most practical telephone systems. In principle, however, the method indicated could readily be expanded to cover a larger number of telephones (and more digits to be dialed) by including more switches of the same type in the switching chain.

Figure 10-10 illustrates another principle which is of fundamental importance in any practical switching system. It will be noted that it shows two trunks between office "A" and office "B" each of which is multipled to selector switches in the two offices. This brings in an additional feature in the switching apparatus under which a switch directed toward a called trunk, but finding it already busy, will automatically move on and hunt for another trunk leading to the same called office. Just as in manual switching, practical economy requires that the numbers of trunks installed between any two offices (or between any other two points in the switching system) be kept at that minimum which (as determined by applications of the laws of probability) will be sufficient to take care of the busy-hour load under reasonably normal conditions. In the large office or exchange area, this may mean that the total number of trunks between two points in the over-all system may run to a considerable number. To keep the number of switching devices to an economical minimum, it is evident that any switch should be able to hunt automatically over the total number of trunks in a group, no matter what its size. Since the standard step-by-step switch can hunt over only ten terminals, graded multiple arrangements are provided when the number of trunks

required per group exceeds ten. With this procedure some of the trunks are accessible to only part of the selectors while the others are accessible to all of the selectors. These trunks are then known as "common" or "individual" trunks. This arrangement increases the efficiency of the trunk groups and, therefore, requires less switches to handle a given amount of traffic.

Two other types of dial switching systems designed for application in large exchange areas are in use in the Bell System. These are known as the **panel** system and the **crossbar** system. Both depend on the principle of first transmitting the dial pulses from the calling subscriber's telephone to a "mechanical operator", usually known as a **sender,** which in turn translates them into other electrical impulses that will operate the necessary switches to establish the through connection. The panel system, although functioning today in many locations, no longer is being installed, so that we may confine our attention in what follows to the crossbar system.

If, as mentioned in Article 9.6, the crossbar switch is thought of as a kind of relay, the crossbar system may be considered as an all-relay system. It employs large numbers of these crossbar switches and of ordinary relays, as well as the multi-contact relays which also were discussed briefly in Article 9.6. Naturally the detailed arrangements are complicated; it will be practicable here to outline only the broad principles of the system's operation. These are illustrated schematically in Figure 10-11, where the through connection from the calling to the called telephone is indicated by the heavy line. As shown, the con-
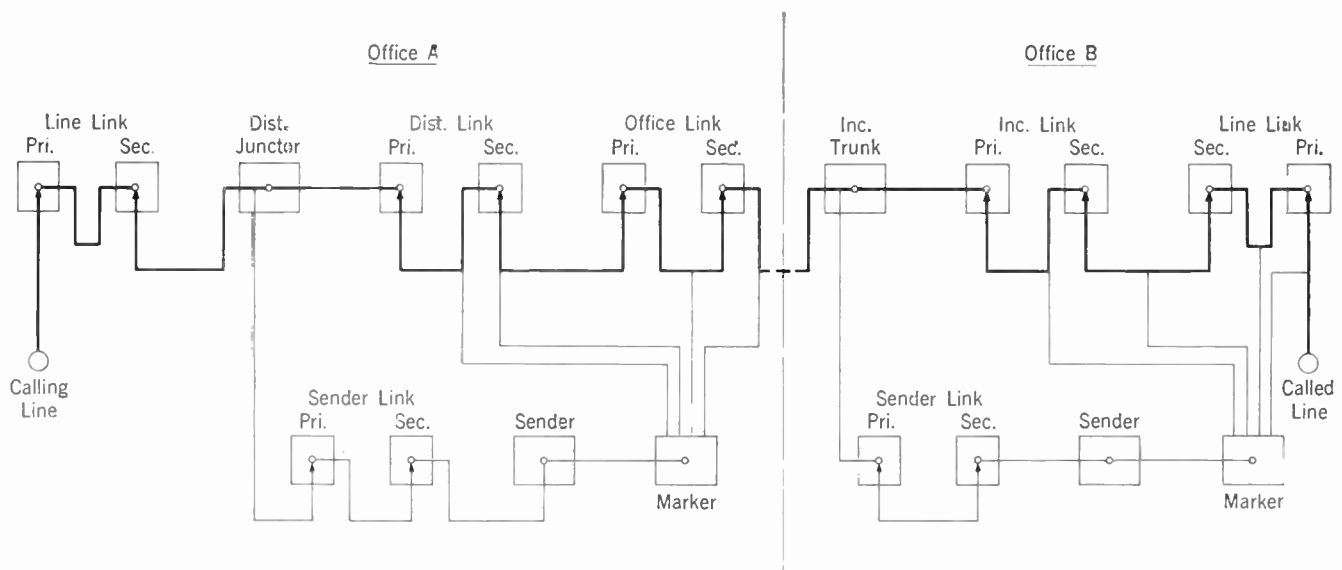
FIG. 10-11 LOCAL CROSSBAR SWITCHING ARRANGEMENT

nection goes through four different switching or link frames, including the line link frame, which acts as a line finder at the calling office and a final selector at the called office. In each one of these frames the connection routes through crossbar switches, generally known as the primary switch and the secondary switch. In the simplest situation where each switch has 100 crosspoints and the switching frame has 10 primary and 10 secondary switches, the arrangement offers the possibility of connecting any one of 100 incoming lines to any one of 100 outgoing lines.

When a subscriber connected to office "A" lifts his handset to make a call, his line is closed to a primary vertical in the line link frame at that office. At the same time relays are operated to call into play a **controller** circuit arrangement (not shown in the diagram) consisting of a number of relays of various types, and associated wiring. This device first definitely locates the calling line on the line link frame and operates the proper selecting and holding magnets to connect the line to an idle trunk or **district junctor** leading to the district link frame. The district junctors extend through a district junctor frame on which is mounted relay equipment that supplies talking battery to the calling line. It also maintains supervision over the call after the connection has been established, times and registers local calls, controls collections or returns of deposited coins at pay stations, etc.

At the same time, the controller hunts for and selects an idle sender and operates selecting and holding magnets in the sender link frame to connect the sender to the calling line. This connection causes dial tone to be sent back to the calling line, indicating that dialing may be started. The dialing pulses are received at the sender and registered there by the operation of relays or crosspoints of a crossbar switch. As soon as the sender has registered the digits indicating the office being called (the first three digits in the case of large multi-exchange areas) the sender operates to locate and connect itself to an idle **marker** through multi-contact relays. It then transmits to the marker information as to the office code dialed, the number of the frame in which the call is originating, the class of service of the calling line, etc. As soon as this information is registered by relays in the marker, the marker first hunts for and selects and "marks" or makes busy an idle outgoing trunk to the distant office. It also selects an idle junctor between the district link frame and the office link frame and then operates the proper selecting and holding magnets in both frames to complete the connection between the selected inter-office trunk and the district junctor to which the calling line had been previously connected. This total operation of the marker is accomplished in a fraction of a second and as soon as it is completed, the marker is automatically released to work on other calls.
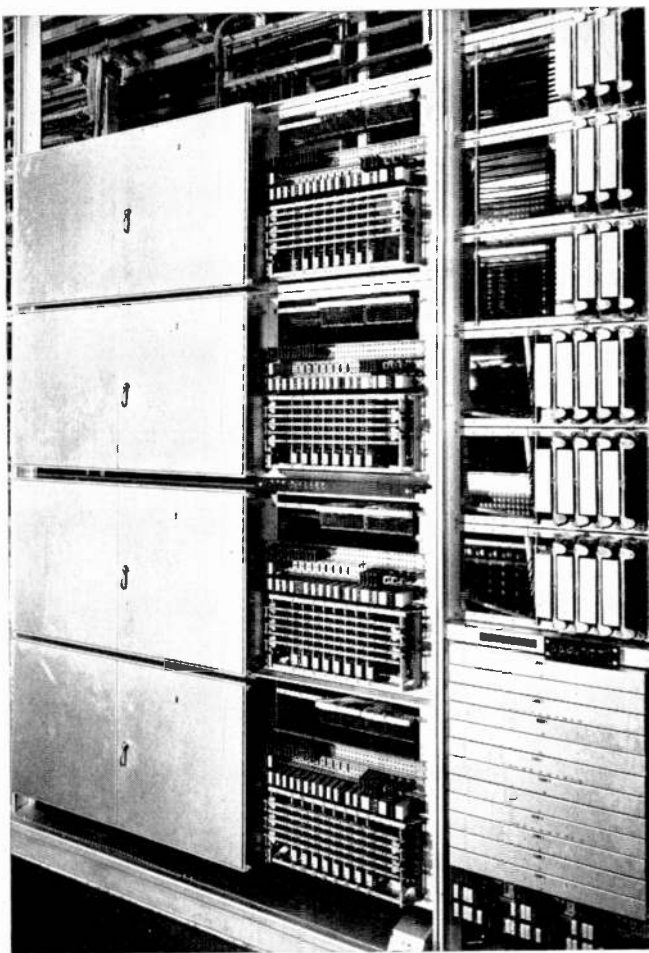
SENDER FRAMES WITH MARKER CONNECTORS AT THE RIGHT

In the meantime, the calling subscriber continues to dial the digits corresponding to the number of the called telephone. These are likewise registered in the sender. As the hundreds digit is dialed, the trunk circuit to the distant office is closed to the incoming trunk frame, which causes a controller at that office to select an idle terminating sender and connect it by the closure of the proper switch-points in the sender link frame. Like the district junctor frame at the calling office, the incoming trunk frame mounts relay equipment which furnishes talking battery to the called subscriber's line, maintains supervision over the terminating end of the connection, furnishes the proper type of ringing current over the called line, and transmits busy tone to the calling line if the called line happens to be busy. After the terminating sender has been connected to the incoming trunk circuit, the called number registered in the originating office sender is transmitted to the terminating office sender where it is

again registered in relays or crosspoints of a crossbar switch. At this point the originating sender having completed its function is disconnected and released for use on another call.

The terminating sender then connects to or "seizes" an idle terminating marker to which it transmits information on the called subscriber number and the number of the incoming trunk frame on which the particular incoming trunk is located. By means of its translator circuit the marker is able to locate the called subscriber's line out of a maximum of about 10,000 such lines terminating at the central office. If this line should be busy, the marker directs the incoming trunk circuit to transmit a busy signal to the calling subscriber. If the called subscriber's line is idle, the marker first makes it busy and then proceeds to test for and select idle links in both the line link and incoming link frames as well as an idle junctor connecting these two frames together. It then operates the proper selecting and holding magnets in the two link frames to operate the necessary crosspoints that will close the circuit through. With this, the incoming trunk circuit transmits ringing current to the called subscriber's line and the terminating sender and marker disconnect and return to normal. As already pointed out, the district junctor at the calling end and the incoming trunk at the called end jointly maintain supervision over the call. When the talking parties disconnect, they function to cause all connections to be restored to normal.

The reader will have noted that in the preceding brief description of a dial call, various apparatus units—particularly the senders and markers—have been spoken of much as if they were persons like manual telephone operators. The functions that they perform are indeed very similar to those performed by operators. Needless to say, therefore, their circuit arrangements are intricate—so much so that their detailed analysis would involve considerably more space than can be allotted to the subject in this book. It should be remembered, however, that they are only machines, having no capacity to exercise either initiative or judgment, but able to do only those things that they were specifically designed to do.

The accompanying photographs give some idea of the physical appearance of a typical sender and a typical marker. The latter may include as many as 1,700 relays. Not all of these operate in establishing any one connection, although it is interesting to note that a typical local dial connection

involves the operation of about 700 relays. Markers and senders are naturally costly pieces of apparatus. As pointed out earlier, however, the marker performs its function in less than a second, which means that it can handle many calls within a short time. Actually it is necessary to provide no more than 5 or 6 markers to handle all the work in one dial system local office. The operating time of the sender is somewhat longer —up to about 15 seconds—so that proportionately a greater number of these apparatus units have to be provided.

10.5 Toll Dial Switching

The establishment of toll or long distance connections by mechanical or dialing methods is not essentially different from the setting up of local connections. As was pointed out in the earlier discussion of manual switching arrangements, long distance switching generally involves a greater number and variety of different situations than does local switching. Thus the long distance connection will frequently require a greater number of switches for completion from terminal to terminal.

In the more usual situation where direct circuits are available between two cities, toll switching equipment is required only at the calling and called points. When direct circuits are not available, however, it may be necessary to route the call through appropriate switching equipment at one or more intermediate points. Also important in toll connections is the fact that when a call extends beyond the area where toll charges are billed as message units, the details of the call must be recorded for billing purposes. These include the number of the calling line, the name of the place called, the called number, and the total time of the call. In **direct distance dialing**, the calling subscriber establishes a long distance connection without the aid of an operator by dialing a seven or ten digit code, depending on whether the called telephone is in his own or a distant numbering plan area. Recording equipment must be provided to register the necessary billing information in this case. In the absence of such equipment, an outward toll operator records the information on a toll ticket, as in full manual operation. She then proceeds to set up the desired connection by **operator distance dialing**, operating keys for the necessary code signals that direct the equipment in her own and other offices to complete the desired connection. In addition to the connection between the incoming calling trunk and an outgoing toll circuit, the overall connection may include one or more through connections at intermediate cities along the route, as well as an inward connection to the local central office in the called city and the loop from that office to the called telephone. The operation of the switching equipment is the same whether directed by an outward operator or by the subscriber's own dial.

Major crossbar dial offices are arranged to select alternate routes automatically in a predetermined order when all circuits on the first-choice route are busy. This important feature, together with the other complexities of toll switching noted above, make it necessary to employ two additional control devices to assist the marker in handling the "brain-work" of the toll switching machine. The first is known as a **decoder** and this is in turn assisted by one or more **translators**. Figure 10-12 illustrates schematically the general arrangement of the complete toll switching machine, including these two devices.

As previously mentioned, the machine is designed to work with either a seven digit or a ten digit code such, for example, as 415-WO6-1748. Here the first three digits designate the distant numbering plan area in which the called city is situated, which may be a State or a part of a State. The second three digits designate the local central office within that area, and the final four the called subscriber's telephone number. When a call is received over an incoming trunk or toll circuit, an idle sender is connected through the sender link frame, just as in the local dial office. As soon as the sender receives and registers the first three digits, it seizes an idle decoder through the decoder connector and gives it this code. The decoder passes this information to a translator which causes it to select the appropriate punched "card" from which it can "read" the routing and other information needed for forwarding the call. It transmits this information to the decoder and drops off. The decoder then seizes an idle marker to which it passes on this information. It also causes the sender to be connected to the same marker. The marker immediately tests the outgoing trunks or lines in the group corresponding to the code information with which it has been supplied, and seizes an idle one, if available. When this is done, it so informs the decoder, which then directs the sender to send all digits forward and leaves the connection. The seizure of the outgoing
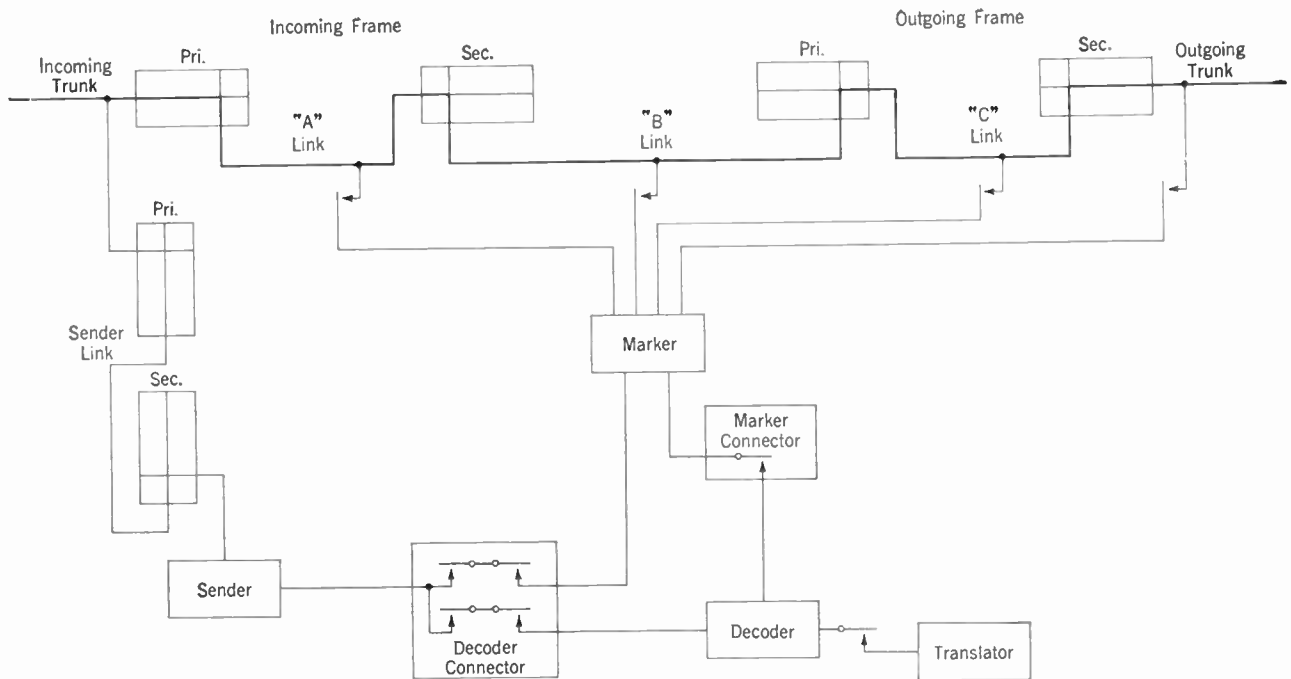
FIG. 10-12 TOLL CROSSBAR SWITCHING ARRANGEMENT

trunk by the marker automatically signals for a sender at the distant point to be connected to the trunk. When this sender is attached, a signal is passed back to the sender at the outgoing office. The latter then forwards the required digits to the distant sender. In the meantime, the marker tests for and selects idle links "A" and "C" in the incoming and outgoing frames and an idle junctor or "B" link between these two frames. It then operates the proper selecting and holding magnets to close the line circuit through the crossbar switches, advises the sender that the connection is completed, and leaves the connection. The sender in turn drops off as soon as this information has been received and it has completed the forwarding of the calling digits to the distant sender. Further progress of the call to another intermediate switch or to its final destination is then under the control of that sender.

The above description deals with the simplest straight-forward connection where a circuit in the first-choice route is available and only the first three digits of the number code require translation. There are many situations when there are two or more routes to or toward a given numbering plan area. In such cases, the translator at the originating point needs to know the exchange code as well as the area code in order to provide the proper routing instructions. Six-digit transla-

tion is required. The translator that is first called in is supplied only with the three digit area code. On reading the card, it recognizes that this information is inadequate and signals the decoder to secure the next three digits from the sender. This translator then disconnects itself. When the decoder receives the next three digits, it calls in another translator to which it now passes the first six digits. From this, the translator is able to select a card that contains the information needed to direct the marker how to proceed to establish the desired connection to or toward the desired central office. With either three digit or six digit translation, the translator supplies the decoder with routing information for not only the first-choice route but for alternate routes as well in their order of priority. If the marker on first instruction finds all circuits busy on the first-choice route, it so informs the decoder which then directs it to search in the first alternate circuit group, and so on until an idle circuit is found. The additional operations required for selecting alternate routes, or for six digit translation, of course increase to some extent the time used in establishing the connection. However, the total time involved is not more than a few seconds. The work time of the individual common control units such as the marker and the decoder is normally less than one second.

[86]

With full dial operation, there is only one outward operator or a machine at the calling end of the connection. It is necessary in either case, for the calling end to know when the called party answers his telephone and when he hangs up. This means that supervisory signals must be automatically transmitted back from the distant end of the connection. The procedures used for this purpose are discussed briefly in the Article following.

10.6 Signaling

In any telephone switching system, it is naturally necessary to provide some means of signaling subscribers and operators. Thus, the subscriber is signaled by the familiar telephone bell which is ordinarily operated by a 20-cycle alternating current supplied from a local central office when an operator presses a ringing key, or when machinery performs a comparable operation. Such a signaling current is not suitable for use over toll circuits of appreciable length since, among other things, it might interfere with telegraph currents being transmitted over the same circuit. This difficulty may be overcome in relatively short toll circuits by using a signaling current of 135-cycle frequency, which is high enough not to interfere with the lower speeds of telegraph services. The necessity for relaying the 135-cycle signal around repeaters, however, limits its usefulness for application on the longer circuits. More reliable signaling on long circuits is obtained by the use of 1000-cycle signaling current. This is a frequency within the normal voice range which suffers no greater losses in transmission than do the voice currents themselves, and it is amplified at intermediate points by the same telephone repeaters that amplify the voice currents. In order to avoid false operation of the signals by voice currents, the signal receiving apparatus is designed to be operated only by a 1000-cycle current interrupted 20 times per second and sustained for at least several tenths of a second. The signals are usually generated by electron tube oscillators. They are in turn detected at the receiving end by appropriate electron tube circuits and converted into direct current or 20-cycle ringing current as needed.

The signaling methods discussed above are known as **ringdown** systems and are employed on "ringdown trunks". They are not suitable for use on circuits arranged for toll dialing operation.

Such circuits must be so equipped that both pulsing and supervisory signals can be transmitted in both directions. Pulsing signals, originally generated by operation of the ordinary dial in the form of dial pulses, may be transmitted over relatively short distances in the form of d-c pulses similar to telegraph signals. In this case, composited or simplexed telegraph channels, separate from the telephone circuits, are used for the transmission. For longer circuits, dial pulses are generally transmitted by interrupting a single frequency tone which is transmitted over the telephone circuit in the same way as the voice signal. This single-frequency signaling method uses a frequency of either 2600 or 2400 cycles on circuits of sufficient bandwidth for such transmission and a frequency of either 1600 or 2000 cycles
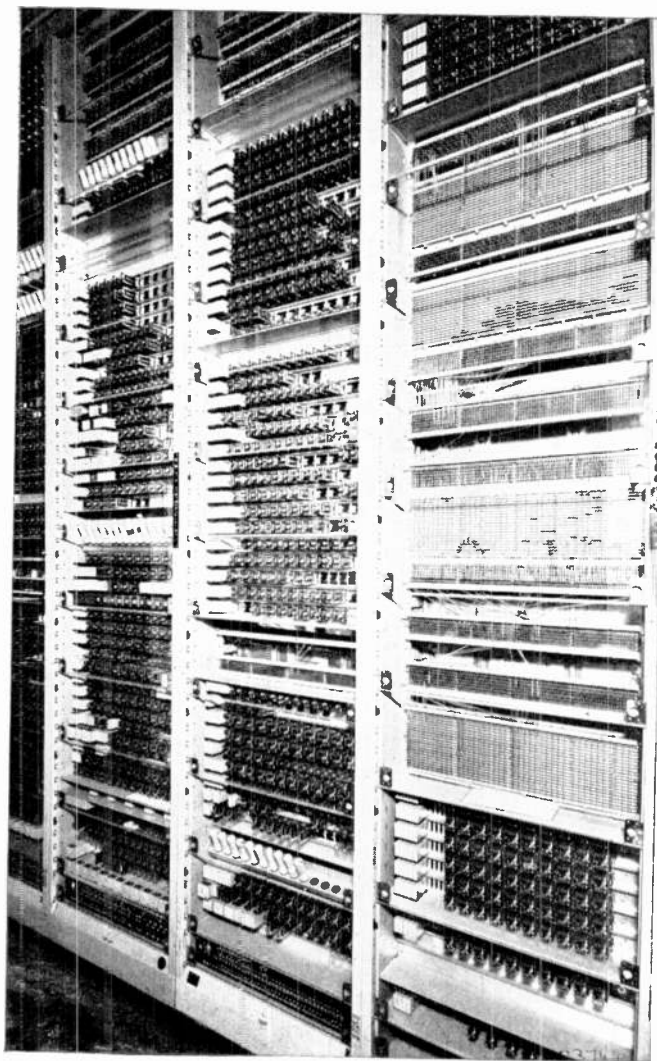
MARKER FRAMES
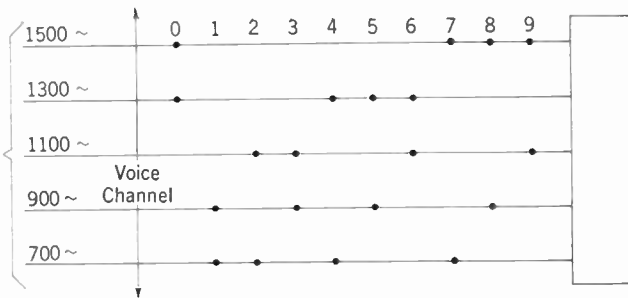
FIG. 10-13 MULTI-FREQUENCY KEY PULSING OF DIGITS

on circuits that cannot satisfactorily transmit the higher frequencies.

Where pulsing signals are generated by key-pulsing, which is much faster than dial pulsing, multi-frequency signals are employed. This generally applies to the longer toll lines. The frequencies used are 700, 900, 1100, 1300, 1500 and 1700 cycles. Each digit is represented by a combination of two of the first five of these frequencies. Figure 10-13 shows the combinations for each of the digits from 0 to 9, inclusive. The 1700-cycle tone, not shown in the Figure, is used in combination with 1100 to provide a priming or "get ready" signal; and in combination with 1500 for a "start" signal.

Supervisory signals are transmitted either in the form of direct current over separate telegraph channels, or by means of a single voice frequency transmitted over the telephone circuit. In the latter case, the supervisory signaling tone should of course not be present when the circuit is being used for conversation, since it is within the voice range. The supervisory signal tone is applied continuously to the line as long as the called telephone is in the "on-hook" condition. When the receiver is taken off the hook, the signaling tone is removed from the line. Its presence or absence causes the operation of relays at the calling end which show when the distant telephone answers or hangs up. These supervisory signal tones must of course be applied to a circuit from both ends in order that a circuit may be used for calls originating at either end. For circuits that are actually or effectively 4-wire, the single frequency of 2600 (or 1600) cycles can be used in both directions. For 2-wire circuits, a 2600 (or 1600) cycle tone is used in one direction and a 2400 (or 2000) cycle tone in the opposite direction.

CHAPTER 11

TELEGRAPH PRINCIPLES

11.1 The Simple D.C. Telegraph Circuit

The earliest device for electrical communication between distant points was called the electric telegraph. This was essentially a direct current system under which pulses of current were transmitted over a single conductor in a code form that could be readily interpreted at the receiving end. The circuit arrangement employed is illustrated by Figure 11-1. Here, when both keys are closed, a d-c circuit may be traced from a battery at the West station through the key and relay at that point to the line wire, thence through the relay and key at the East station, and back through the earth to the battery. The relays at each end in
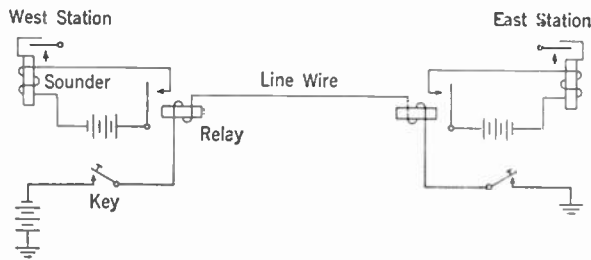

FIG. 11-1 ELEMENTARY TELEGRAPH CIRCUIT

turn control local circuits which include a separate battery and a sounder, a relay-like instrument that produces audible clicks when its armature moves from one position to the other. Opening and closing the key at one end, while the key at the other end is closed, causes both sounders to operate accordingly. Thus, two skilled operators who understand the code being employed, may communicate over the line. Manual operation of the key at the West station produces signal clicks at the East sounder, which the East operator can interpret. The West operator also hears the same signals from his own sounder. If the East operator wishes to send, he opens his own key, which silences both sounders. The West operator, recognizing the "break", then closes his key and listens for signals sent from the East key.

It is hardly necessary to state that modern telegraph practice employs circuits and circuit ar-

rangements of a much more sophisticated nature than the elementary circuit described above. Nevertheless, it is believed that an understanding of current telegraph techniques, may be best approached through study of some of the basic direct-current telegraph circuits. Some of the principles involved will also apply to other kinds of d-c signaling systems, including the transmission of relatively slow dial pulse and supervisory signals and the extremely rapid signals of the pulse code modulation systems discussed in Chapter 27. At this point, however, attention will be confined to what are essentially direct current phenomena. Alternating current applications in telegraph systems are necessarily postponed to later Chapters.

11.2 Telegraph Transmission Facilities

The great majority of long-haul telegraph circuits in the Bell System are now obtained through the use of "voice-frequency carrier systems" by means of which up to 18 two-way telegraph circuits are derived from a transmission facility that would otherwise be capable of use as a single telephone circuit. Before the advent of such carrier systems, practically all long telegraph circuits were derived from open wire or cable wire conductors that were used simultaneously for telephone transmission. Naturally, this required the employment of some device that would sep-
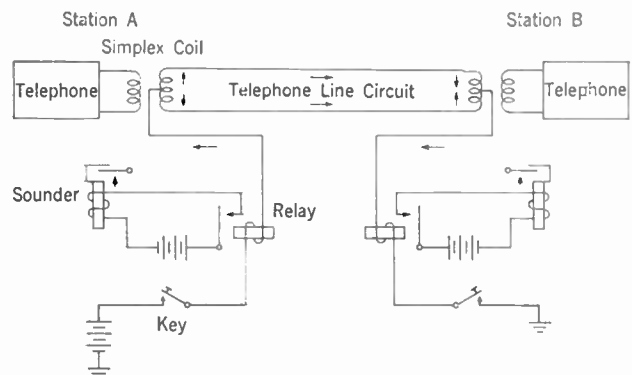

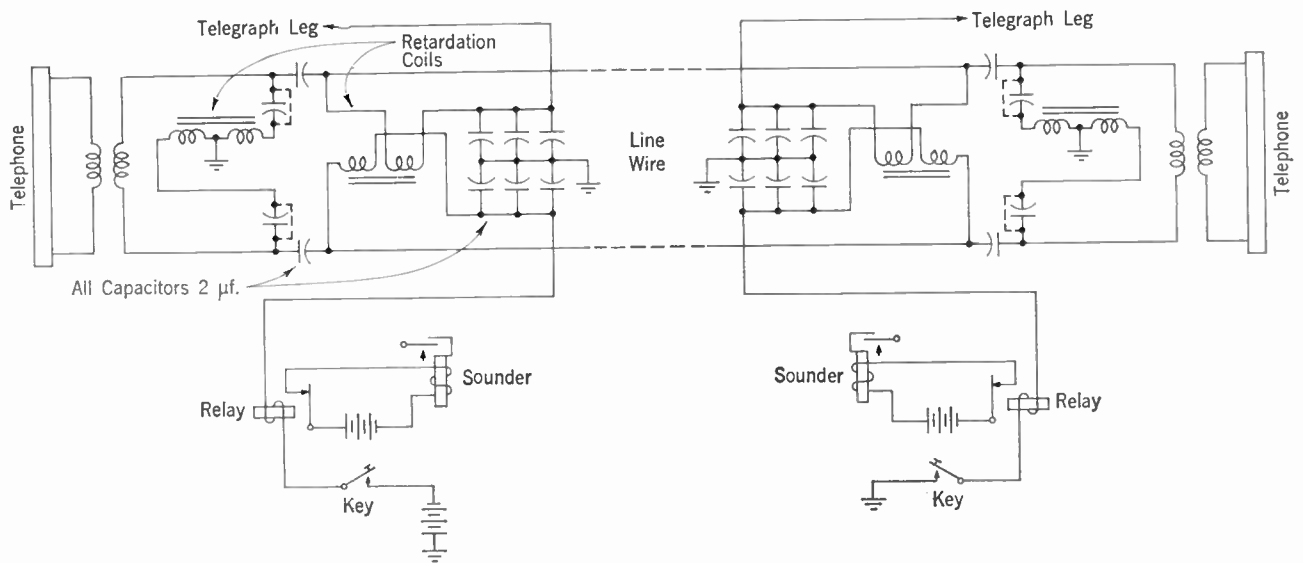FIG. 11-2 TELEGRAPH CIRCUIT ON SIMPLEXED TELEPHONE CIRCUIT

FIG. 11-3 TELEGRAPH CIRCUIT ON COMPOSITED TELEPHONE CIRCUIT

arate the telephone currents from the telegraph currents at each line terminal.

The oldest device serving this purpose is the **simplex set** by means of which one grounded telegraph circuit may be obtained from the two wires of a telephone circuit. Its principle is illustrated by Figure 11-2. As indicated there, the "telegraph legs" are connected to the midpoint of the line windings of simplex or "repeating" coils which are inserted in the telephone circuit at each terminal. For reasons outlined in Chapter 14, these coils do not interfere with the alternating current telephone transmission. The direct telegraph current divides in halves which flow in the same direction through the two line wires, as indicated by the arrows. Any change in current value at the "make" or "break" of the telegraph key then has no effect on the telephone circuit because the magnetic field established by the telegraph current in one half of the coil winding is neutralized by the equal but opposite field set up in the other half of the same winding. Satisfactory operation requires, however, that the two line conductors have identical electrical characteristics, including not only equal or balanced series resistance but also equal capacitances and leakages to other conductors and to ground. Otherwise, the telegraph currents will not divide into exactly equal parts at the midpoints of the line windings and the difference will induce currents in the "drop" windings of the coils which may be heard as clicks or "thumps" on the telephone circuit.

Differing radically in principle from the simplex

set, the **composite set,** illustrated by Figure 11-3, permits a grounded telegraph circuit to be derived from each of the two wires of a telephone circuit. Superficially, this device may be considered as an electrical filter designed to separate the direct telegraph currents from the alternating telephone currents. The $2\mu f$ capacitors inserted in each line wire between the telephone drop and the point where the telegraph legs are taken off prevent any steady direct currents from reaching the repeating coils while readily transmitting the alternating telephone current. By themselves, however, they could not protect the telephone circuits from clicks resulting from the sudden voltage changes that occur when the telegraph key is opened or closed. The composite set includes a number of retardation coils and capacitors which serve to minimize these transient effects by reducing the abruptness of the direct current changes at the make or break of the key. Thus, the retardation coils in series with each telegraph leg insert considerable inductance which opposes a sudden building up of the current at the make of the key and retards its rate of decay when the key is opened. The three $2\mu f$ capacitors connected in parallel to ground assist the inductance by storing a small quantity of electricity while the key is closed and discharging this through the inductance when the key is opened. As a further precaution, a bridge may be inserted across the telephone circuit on the drop side of the series capacitors, as shown in the drawing. This provides a path for any transient charges that may

[90]

appear on the drop side of the series capacitors to "leak" to ground.

In current telegraph practice, simplex and composite sets are used chiefly for deriving relatively short telegraph circuits such as branches or legs extending from a central point on a main long-haul route to outlying towns or cities. Both types of sets are also extensively used in connection with machine switching systems to derive direct current circuits for the transmission of dial pulses and supervisory signals.

11.3 Telegraph Codes and Signals

In the simple manually operated telegraph systems discussed thus far, the elements of the signal code consist of "marks" and "spaces". Current flowing in the line represents a marking signal; absence of current represents a spacing signal. Systems operating in this way are said to employ neutral transmission. For the longer circuits, there are usually advantages in employing polar transmission, in which the marking signal is represented by direct current flow in one direction and the spacing signal by equal current flow in the opposite direction.

Manual telegraphy, which is now largely of historical interest, used in this country what was known as the Morse code. In this, the letters of the alphabet and the numerical digits were represented by different combinations of varying numbers of short and long marking signals (dots and dashes) separated by short or long spacing signals. Where machines are used to transmit and receive telegraph signals, the code usually employs the same number of units, spacing or marking, to represent each letter or digit and the signal units are all of the same length. The code so employed in ordinary teletypewriter practice is shown in Figure 11-4. This uses five units or elements to provide a maximum of 32 possible signal combinations. With the employment of upper and lower case positions as in the ordinary typewriter, these are sufficient to provide for the letters of the alphabet, the numerals, and several miscellaneous symbols in common use, as well as for necessary machine operations such as spacing, shifting, line feed, and carriage return.

For other purposes, codes made up of a larger number of signal units may be necessary. This is true of teletypesetter operation which requires a 6-unit code to enable the transmission of both upper and lower case letters as well as numerals

and other symbols. Pulse code modulation systems may use 6-unit or 7-unit codes to provide maximums of 64 or 128 possible signal combinations.

11.4 Principles of Teletypewriter Operation

The essential principles involved in the transmission of telegraph-type signals by machine methods may perhaps be most easily understood from an analysis of the operation of an early teletypewriter device, known as a distributor. This consisted of two circular commutator faces over each of which a motor-driven brush arm may rotate. Figure 11-5 shows schematically the send-
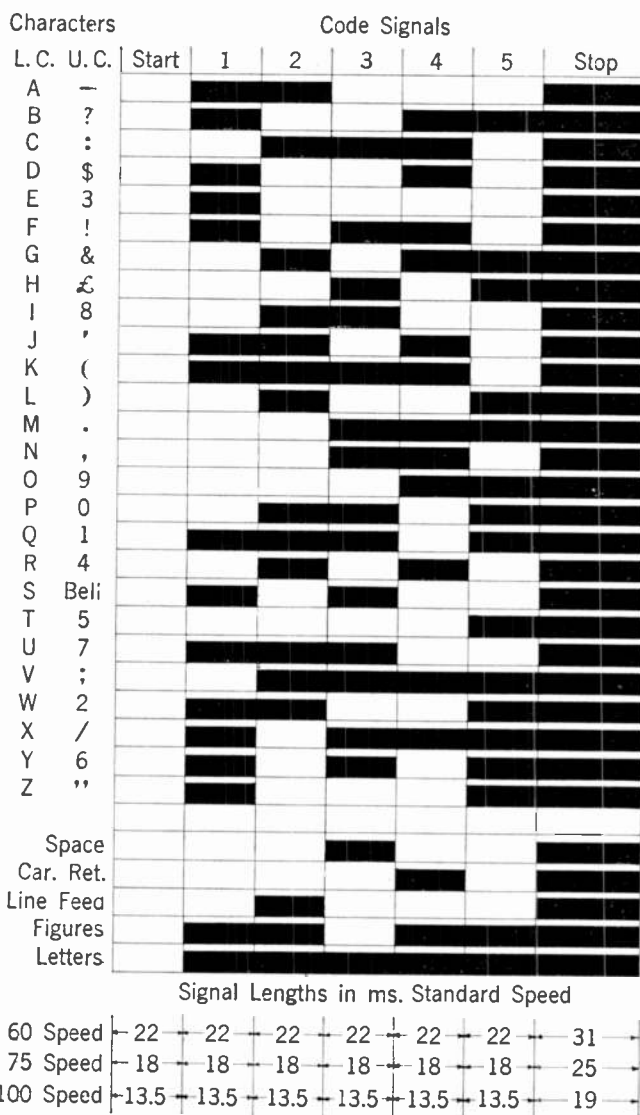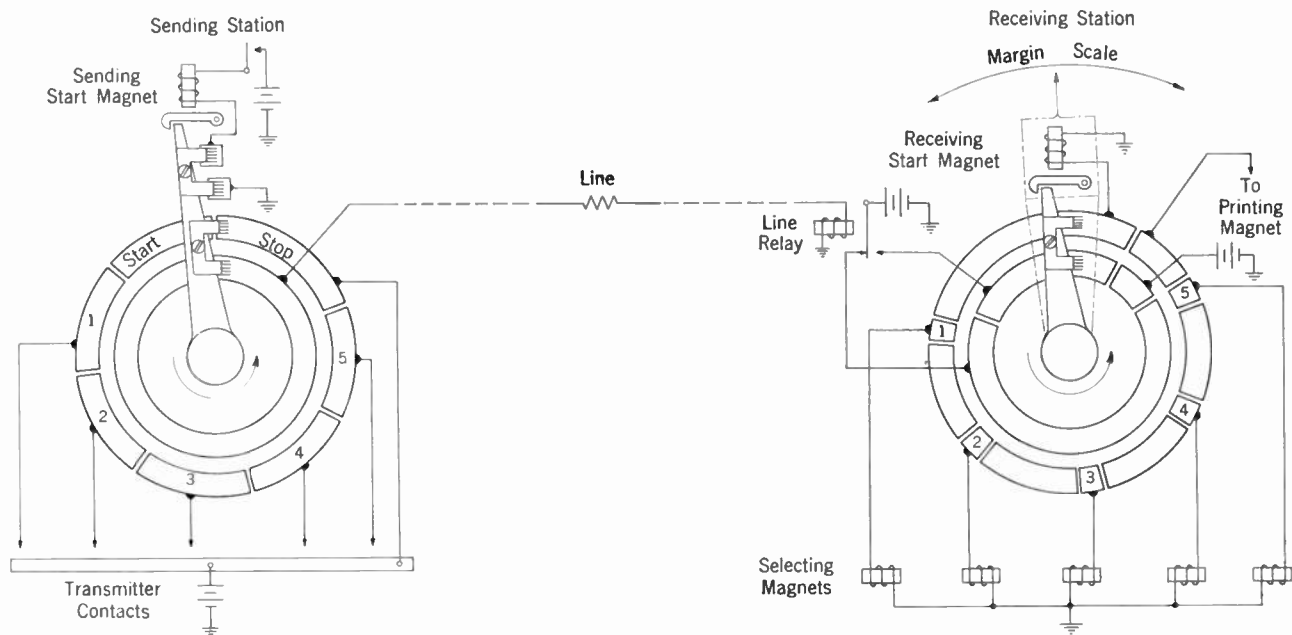
FIG. 11-4 TELETYPEWRITER CODE

FIG. 11-5 THEORY OF START-STOP TELETYPEWRITER SYSTEM

ing face of such a distributor at one end of a telegraph line and the receiving face of a like distributor at the other end of the line. The five numbered segments in the outer ring of each commutator face are electrically connected respectively to sending and receiving teletypewriter machines. These are mechanical devices of various designs, which will not be described here. It will be sufficient for our purpose to understand that the sending teletypewriter applies battery to the appropriate leads to the sending commutator when the key for a given character is operated; and that the receiving teletypewriter automatically prints the character corresponding to a given combination of operated selecting magnets at the receiving commutator.

For accurate reception of signals, it is of course necessary that the receiving device operate in synchronism with the sending device. In most teletypewriter systems, this is accomplished by momentarily stopping the machines after the transmission of each character and starting them again together at the beginning of the next character. This is the purpose of the Start and Stop signal pulses included in Figure 11-4. This method of operation, as well as the other details of signal transmission, may be followed by referring again to Figure 11-5.

At the beginning, the brush arms of both distributor faces, which are coupled by friction

clutches to the driving shafts of motors running continuously at approximately the same speed, are held stationary by the latches of the sending and receiving start magnets. The operation of any character key of the sending teletypewriter closes a circuit to ground through the winding of the sending start magnet and the outer pair of brushes. This pulls up the latch momentarily and allows the brush arm to start to rotate. As the inner pair of brushes moves off from the stop segment and on to the start segment, battery connection to the line is broken and a spacing signal is transmitted. This "start-pulse" releases the receiving line relay which connects battery through the pair of brushes to the receiving start magnet. The resultant operation of its latch then allows the receiving brush arm to start to rotate also.

Operation of a key of the sending teletypewriter will also have caused battery to be connected to the appropriate leads to the numbered segments of the sending commutator. Assuming, for example, that the key for the letter, D, was operated, battery is connected to segments 1 and 4 in accordance with the code for D which is mark, space, space, mark, space. Then, when the sending brush arm moves off from the start segment and on to segment 1, the line is closed to battery and the receiving line relay is operated to connect battery to the large inner segment of the receiving distributor face. The rotating receiving brush

arm thus connects battery to the shorter segment 1 of the receiving face while the sending brush arm is still passing over segment 1 of the sending face. Selecting magnet No. 1 is energized. As the rotation continues, selecting magnets 2 and 3 are not energized because no battery is connected to segments 2 and 3 of the sending face. Selecting magnet No. 4 is operated, however, while No. 5 is not. At this point, the mechanisms of the receiving teletypewriter are properly placed to print the letter, D, as soon as the receiving brush arm passes on from segment 5 and connects battery to the printing magnet. In the meantime, the sending brush arm has passed on to its stop segment, thereby transmitting a long marking signal to the line. Its further rotation is then stopped by the start magnet latch which was released as soon as the outer pair of brushes opened the circuit through the winding at the beginning of the operation. The received stop signal holds the receiving line relay closed so that the receiving start magnet is not operated, and the receiving brush arm is also stopped by its latch as it completes the revolution. Both distributors are then in position to handle the next character.

As mentioned earlier, signal distributors of the kind described above have been generally superseded by somewhat more sophisticated mechanical devices. The principle of operation remains the same, however, and is applied not only to telegraph circuits but to other direct current signaling systems, including pulse code modulation systems which must operate at relatively fantastic speeds.

11.5 Telegraph Repeaters

As in all other electrical communication systems, telegraph signals are subject to attenuation or loss of energy when transmitted over any considerable distance. These losses must be counteracted by the use of appropriate devices at intervals along the circuits, which will restore the signals to their original strength. For long distance sections where, as has already been stated, telegraph transmission usually employs carrier methods similar to those of telephony, electronic amplifiers or repeaters are used to accomplish this purpose. Such devices, and their applications in both telegraph and telephone transmission systems, are discussed at some length in later Chapters.

While most long-haul telegraph transmission

thus fails in the alternating current category, the telegraph signals at their ultimate terminal, whether sending or receiving, are direct current. Transmission over the loops extending from the telegraph central office to the customer's premises is usually by direct current. This may also be true for relatively short inter-office lines connecting a major office to tributary offices at outlying points. In either case, it is necessary to employ some kind of direct current repeating device at the central offices to counteract the d-c line or loop losses. There are many kinds of such devices and all depend essentially upon operating a relay by the weakened signal pulses to connect fresh battery to the line. The basic principles may be followed by analyzing the operation of a widely used telegraph repeater known as a differential duplex set.

This repeater may be arranged to handle full
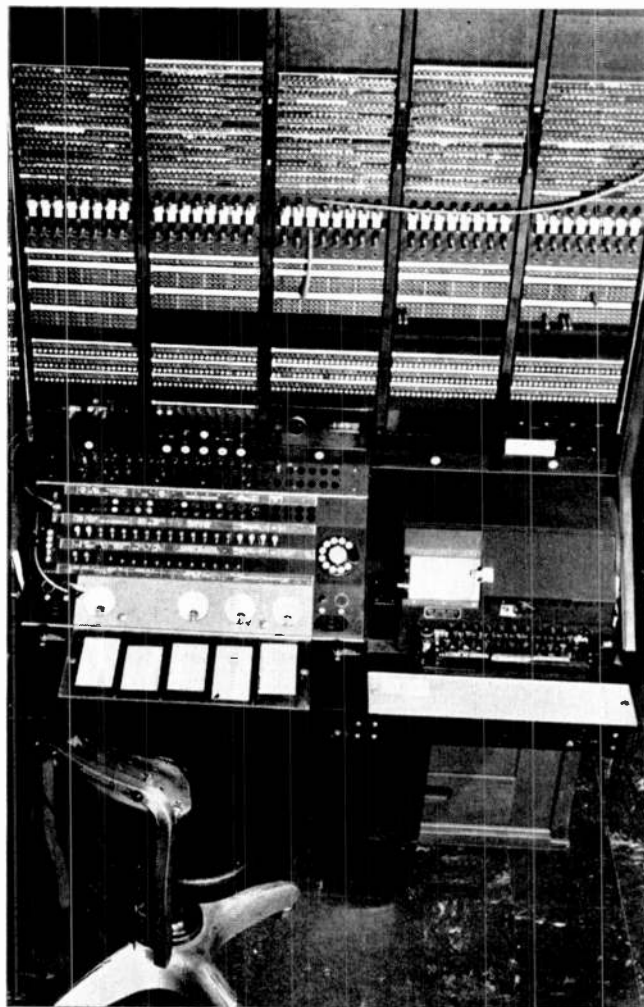
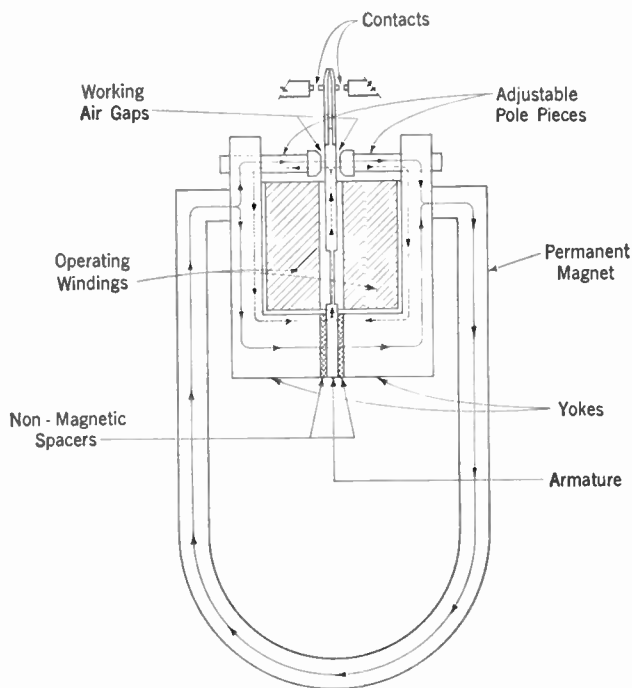TELEGRAPH SERVICE BOARD POSITION

[93]

FIG. 11-6 MAGNETIC CIRCUIT OF THE DIFFERENTIAL POLAR
RELAY

duplex service, in which signals are transmitted simultaneously in both directions; or **half duplex** service, in which signals are transmitted alternately in either direction. It is designed to operate with polar signals since these are generally preferred in modern telegraph practice for reasons that are explained in Article 11.6 following. The repeater includes polar relays of the type illustrated in Figure 11-6. This relay employs a differential principle which depends upon winding the magnetic core with two equal but opposite windings. If equal currents flow in the same direction through both windings, the magnetic field produced by one winding is exactly neutralized by that set up by the other winding. The relay is polarized by means of the permanent magnet and the split magnetic circuit so that its armature operates toward one contact or the other depending upon the direction of current flow in the dominant winding.

Figure 11-7 shows the essential elements of a differential duplex set provided at a terminal telegraph office to repeat signals between the line and the local loops to a customer's office. In this case the set is arranged for full duplex service and there are two loops, since the customer may be sending and receiving at the same time. To fol-

low the repeater's operation, note that the two windings of the receiving relay are connected on one side respectively to the real line and to a balancing network called an artificial line. The other ends of the windings are connected together to the armature of the sending relay. When the artificial line is adjusted so as to exactly balance the real line electrically, current coming from either battery connected to the contacts of the sending relay divides equally between the two windings. Since these are connected differentially, the resultant magnetic flux is zero and the relay is not operated. Current coming from the line, on the other hand, flows through the two windings of the receiving relay in series. This produces aiding magnetic fields and causes the relay armature to move to one or the other of its contacts, depending upon the polarity of the incoming current. Thus, there is no interference between the sending and receiving circuits and the two can be operated quite independently of one another.

Operation of the sending relay is of course under the control of the customer's instruments in the sending loop. The upper winding of the sending relay is known as the operating winding and the lower as the biasing winding. When the sending loop key is closed, the magnetic fields produced by the two windings are in opposition because current is flowing in the same direction through each. However, the lesser applied voltage together with the resistance R, limits the current in the biasing winding to a substantially smaller value than that in the operating winding. The relay armature is therefore held on the marking contact. When the loop key is opened, only the biasing winding is effective and the armature is operated to the spacing contact.
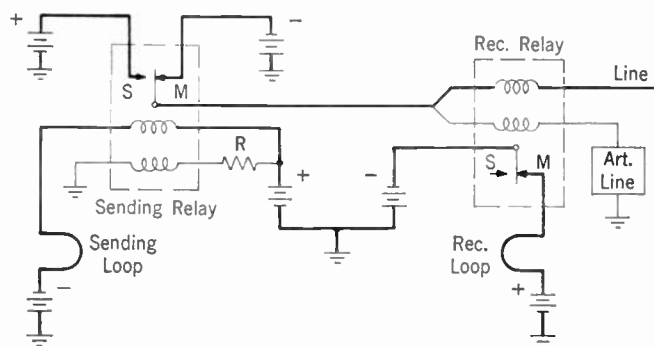

FIG. 11-7 TERMINAL DIFFERENTIAL DUPLEX SET
ARRANGED FOR FULL DUPLEX SERVICE

[94]

The differential duplex set, arranged for half-duplex operation at a circuit terminal, is illustrated schematically in Figure 11-8. In this case only one loop to the customer is required since signals are to be transmitted in the two directions only on an alternate basis. But the set must then be so arranged that the operator at either end of the line, when receiving, can interrupt or break the operator sending at the other end. An additional polar relay, known as the break relay, is provided for this purpose. Also, to permit sending and receiving over the same loop, the armature of the receiving relay is connected to the apex of the sending relay windings and an artificial line or network R is added to balance the electrical characteristics of the loop.

With these connections, the sending operation is essentially no different from that of the full-duplex set of Figure 11-7. When receiving, the operation of the receiving relay armature impresses alternate positive and negative signals on the loop, provided that the break relay armature remains in the marking position. Analysis of the current flow in the windings of the sending and break relays shows that the armatures of both of these relays will remain on their marking contacts as long as the loop is closed. Thus, when the receiving relay armature is in the marking position, the positive battery of the loop is reinforced by the negative battery connected to the marking contact of the receiving relay. A relatively large current then flows in the operating windings of the break and sending relays. At the same time, a current of half this value is flowing in the same direction through the biasing windings of both relays. This opposes the effect of the current in the operating windings but, since the latter is much larger, it is controlling and the relay armatures are held in the marking position. When the receiving relay armature moves to the spacing contact, positive battery at the marking contact of the break relay counteracts the positive battery of the loop so that no current flows in the operating windings of the sending and break relays. At the same time, the direction of current flow in the biasing windings is reversed. The armatures of both relays are now held in their marking position by the reversed current in the biasing windings.

The purpose of the break relay is to assure a rapid and positive break when the loop is opened. If it were not included and the positive and negative batteries were connected directly to the spacing and marking contacts of the receiving relay

instead of as shown in the drawing, the operation of the receiving relay while the loop was open would cause the sending relay to transmit inverted signals back over the line. This would not result in the most effective breaking action. With the break relay in the circuit, opening the loop causes its armature to move to the spacing contact when the receiving relay is marking because there is then no current in the operating winding. This results in negative battery being connected to both contacts of the receiving relay, which compels the latter to transmit a marking signal to the sending relay windings. The resultant current in the biasing winding of the sending relay causes
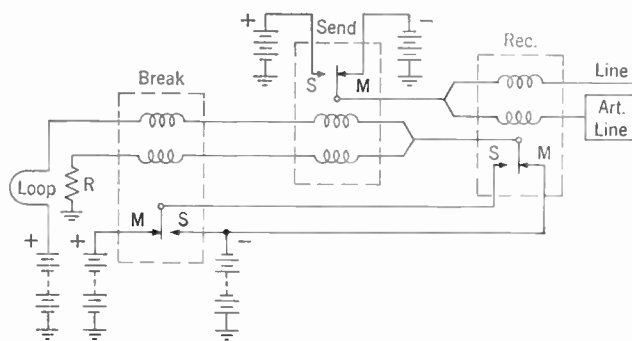
its armature to move to the spacing contact. This sends a continuous spacing signal over the line to stop the sender at the distant end.

11.6 Telegraph Transmission

Ideally, the direct current telegraph signals received over a transmission line should be exactly the same as the signals sent. Thus, we may assume that the Morse code signal for the letter A, made by opening and closing a telegraph key, has the "wave shape" shown in Figure 11-9(A). If the transmission line is purely resistive, the received signal will have the same wave shape except for some decrease in the current strength due to IR drop in the line. For very short loops, or connections within a central office, this situation may be closely approximated. Loops more than two or three miles in length and practically all line circuits, however, contain enough inductance and capacitance to affect the signal wave shape materially. As pointed out in Chapter 8, the pres-

[95]

FIGURE 11-9

ence of inductance causes the current at the closing of the key to build up from zero to maximum value on a sloping curve, somewhat as shown in Figure 11-9(B). It also slows down the falling off of the current when the key is opened. Capacitance in the circuit tends to reshape the wave further, as indicated by the shaded portions in Figure 11-9(C). Insertion of inductors and capacitors, as in composite sets or similar apparatus, naturally increases the extent of the wave shaping still more.
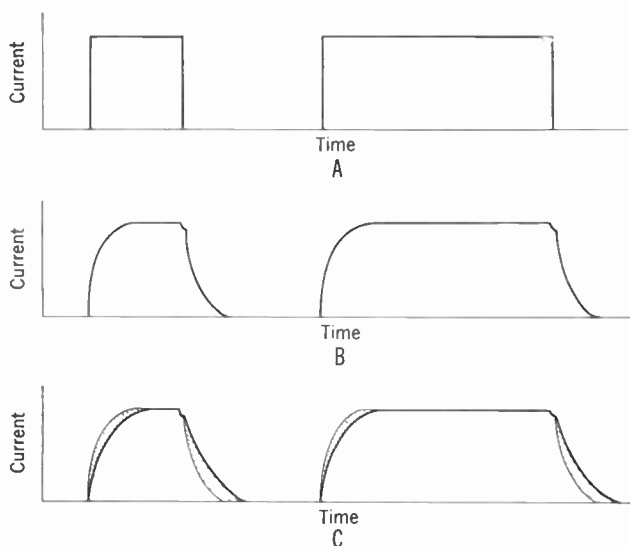
The basic significance of the signal wave shape, from the viewpoint of telegraph transmission, may be readily understood by considering its effect on signal reception with a simple neutral relay. Such a relay may be adjusted to operate at any one of a considerable range of current values and to release at corresponding lesser values of current. If a received signal wave was essentially "square," as in Figure 11-9(A), the relay would operate and release at the same time regardless of its adjustment. When the wave is not square, as is usually the case, the time length of the received signal is definitely affected by the relay adjustment. Thus, in Figure 11-10(A) the relay is adjusted to operate at time O and to release at time R; the length of the received signal is then represented by T. If now the adjustments are changed to permit operation at O_1 and release at R_1, as in Figure 11-10(B), the length of the signal is increased to T_1. In telegraph parlance, the signal is now "heavy". Adjustment of the relay

in the opposite direction would have shortened the signal, or made it "light". These two conditions are called **marking bias** and **spacing bias** respectively. Either type, if of sufficient magnitude, may make it difficult to "read" the received signals accurately.

The effect of signal wave shape on bias may be more critically analyzed by considering neutral transmission employing polar relays. The armature of the polar relay is held in the spacing position by a constant current in its biasing winding instead of by a spring, as in the neutral relay. Normally the biasing current is set at a value equal to one-half of the full line current when marking. Thus in Figure 11-11, which again represents a wave shape for the Morse A, the full line current is 60 milliamperes and the current in the biasing winding is 30 ma. If the relay pole-pieces are properly adjusted so that both marking and spacing poles are equidistant from the armature, the armature is free to move to its marking position as soon as the current in the line winding reaches a value slightly higher than 30 ma. Similarly, the armature returns to the spacing contact as soon as the line current decreases slightly below 30 ma.

It will be evident that when the signal wave shape is curved as shown, any change from normal in the biasing winding current value, in the full line current value, or in the pole-piece adjustments, may change the length of the received signals. A lowered value of relay bias current, for example, causes the operating and release points to move downward on the signal curve with resultant marking bias. The same effect appears if the marking pole is adjusted too close to the armature, or the spacing pole too far away. Poor adjustment of the pole-pieces in the opposite direction, or too high value of relay biasing current, on the other hand, causes light signals or spacing bias. For the same reasons, higher than normal
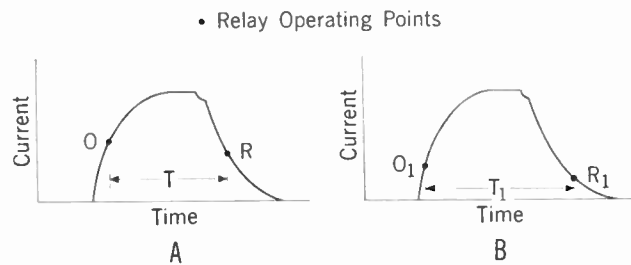

• Relay Operating Points

FIG. 11-10 EFFECT OF RELAY ADJUSTMENT ON TELEGRAPH SIGNAL LENGTHS

line current produces marking bias while low line current results in spacing bias.

Either marking or spacing bias may also be caused by the characteristics of the transmission circuit, even though all relay adjustments and current values are normal. This results from different curvature in the leading and trailing edges of the signal pulse. Referring again to Figure 11-11, it will be noted that there is a definite interval between the time the circuit is closed and the time the relay operates because of the curvature of the rising current wave. This interval, which may range from a fraction of a millisecond to several milliseconds, is called a **space-to-mark transition delay** and abbreviated S-MTD. Likewise, when the circuit is opened, there is a mark-to-space transition delay (M-STD) as the current falls off from maximum value to the relay release point. The length of each delay is determined solely by the shape of the current wave. If the two delays are equal in length, the signal length is not affected and the circuit introduces no bias. If the M-STD is greater than the S-MTD, all marks are lengthened accordingly and all spaces shortened a like amount. The circuit thus causes marking or positive bias. Contrariwise, if the S-MTD is longer than the M-STD, all spaces are lengthened and marks shortened to produce spacing or negative bias.

For a given signal, such as that shown in Figure 11-11, it is of course possible to measure the transition delay time or the extent of bias in terms of milliseconds. Because several different signaling speeds are employed in telegraph practice, however, it is generally more useful to measure bias as a percentage of the unit signal element or pulse. Thus, the unit pulse in 60 speed teletypewriter operation is 22 ms in duration, as indicated in Figure 11-4. A positive bias of 20% in this case would lengthen marking signals by 4.4 ms. If transmission were at 100 speed where the unit signal length is 13.5 ms, the same percentage bias would lengthen the signals by 2.7 ms.
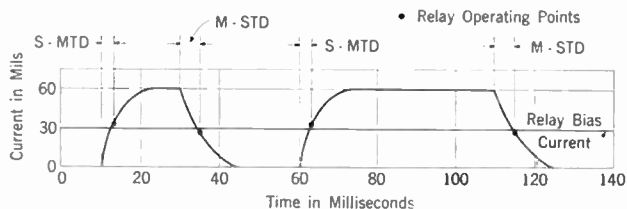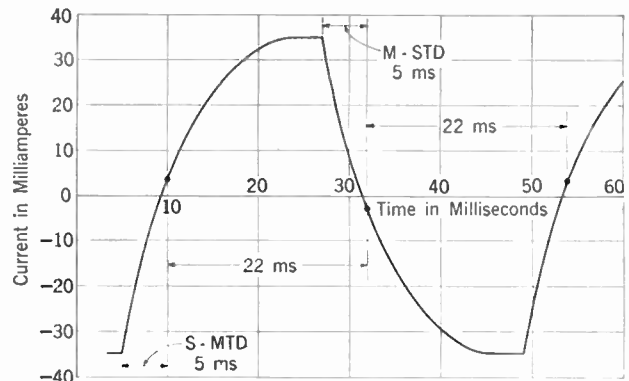
FIGURE 11-11

FIG. 11-12 SIGNAL WAVESHAPES IN POLAR TELEGRAPH CIRCUIT

With true polar transmission, the probability of bias distortion of signals is substantially reduced. The ideal wave shape for such operation is illustrated in Figure 11-12. Here, as would normally be expected, the marking and spacing waves are identical in form. If, in addition, the positive and negative line currents are equal in value and all relays are adjusted to operate and release at approximately the middle of the wave, the mark-to-space transition delays must always be exactly the same as the space-to-mark delays. Under these conditions, transmission is completely free from bias. It is necessary to note, however, that relays can get out of adjustment as readily with polar as with neutral transmission. Also, it is possible for conditions to exist, such as a difference in ground potential between the terminals of a circuit, that cause the marking and spacing line currents to have unequal values. Either effect will unbalance the transition delays and thus produce bias.

It is important to keep in mind that, in either neutral or polar operation, any bias present is a constant whose value is determined by the characteristics of the line circuit and of the relays and other equipment associated with it. Furthermore, it affects all signals, regardless of their length, in exactly the same way and in exactly the same amount. Bias is always undesirable and, if severe enough, can cause false operation of receiving equipment. Unless the curvature of the wave shape is too extreme, however, it can usually be kept within satisfactory working limits by appropriate relay adjustments.

In the discussion thus far, it has been tacitly assumed that all marking or spacing signals were of sufficient length to permit the line current to reach its full or steady state value before the start
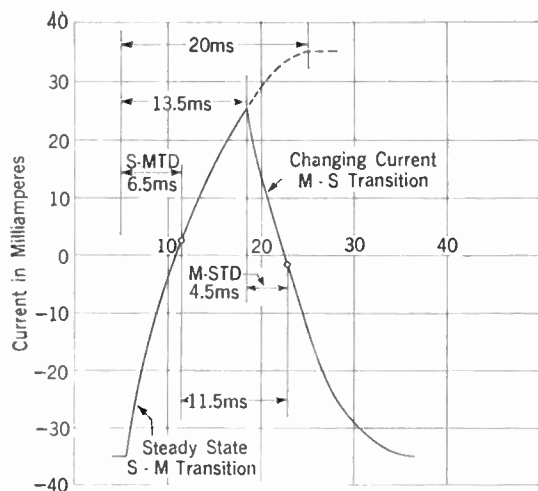
| 97 |

FIGURE 11-13

of the next transition. This is not necessarily true in high-speed telegraph transmission. Referring to Figure 11-4, it may be noted that the shortest marking and spacing signals in 100-speed teletypewriter operation occupy a time interval of only 13.5 ms and that the majority of the character codes include one or more of such short marks or spaces. Now note that in a circuit having the wave shape pictured in Figure 11-12, the current takes approximately 18 ms to rise from its steady state spacing value to its full marking value. A 13.5 ms signal would therefore encounter a transition while the current was still changing and before it reached its full value. This introduces another kind of signal distortion, known as **negative characteristic distortion**. Its effect is to shorten all short signals, whether spacing or marking; it has no effect on the longer spacing or marking signals which form a part of most of the codes.

Examination of Figure 11-13 will perhaps help to clarify this situation. Here a wave shape is assumed in which a total of 20 ms is required for the current to change from a steady state spacing value of −35 ma to a steady state marking value of +35 ma. If a marking signal only 13.5 ms long is applied to a circuit having this wave shape, the rising current has time to reach a value of only about 25 ma before the signal is reversed to spacing. The receiving relay operates to marking as the current passes through zero value, giving a S-MTD of 6.5 ms. But since the reversed spacing current now starts from less than full marking value, it takes less time for it to reach zero value where the receiving relay operates to spacing.

The M-STD is only 4.5 ms. The net result is to shorten the received marking signal by 2 ms to a value of 11.5 ms. Since the positive and negative wave shapes are normally identical in form in polar operation, it is obvious that a short spacing signal would be affected in exactly the same way. The transition from space to mark would likewise occur while the current was still changing and the received spacing signal would be shortened accordingly.

Positive characteristic distortion is also possible. This occurs when the electrical characteristics of a transmission line are such as to produce a transient oscillation that causes the changing current of a transition to increase momentarily to a value greater than the normal steady state value. This may increase the time required for the current to return to zero value enough to lengthen all short signals.

Whether either type of characteristic distortion is present, and its extent, is determined by the speed of transmission and the consequent length of the shortest signals in relation to the wave shape, which is in turn determined by the fixed characteristics of the circuit. Because it affects only short signals, it is also interesting to note that it has no effect whatever on those code characters which are made up only of longer signals.

In telegraph transmission, the practical importance of bias or characteristic distortion depends upon the amount by which signals are affected in relation to their normal values—i.e., the percentage changes in the signal lengths. In a relatively short circuit having repeaters only at its terminals, it is evident that distortion must be rather severe to cause false signal reception. Such a re-
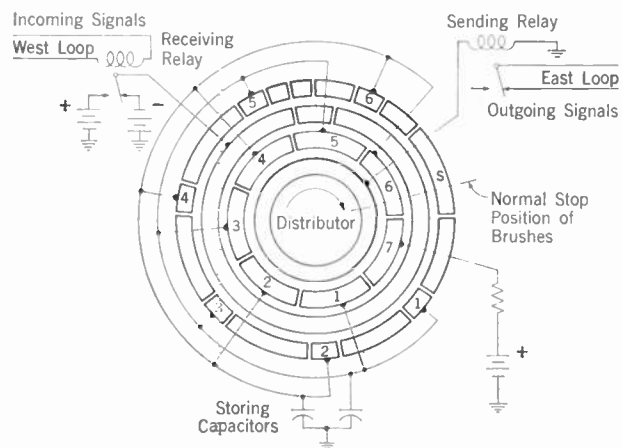

Fig. 11-14 Principle of the Regenerative Repeater

[98]

sult is further guarded against in teletypewriter operation by making the signal receiving elements on the receiving face of the distributor much shorter in length than the corresponding elements on the transmitting face. Thus, referring again to Figure 11-5, it may be noted that the numbered receiving segments are only about 20% as long as the transmitting segments. If the machines are so adjusted that the receiving segments are passed over at the center point of an undistorted incoming signal, any incoming signal would register unless it was distorted by more than 50% in one direction or the other. Distortion of such magnitude as this is unlikely in short telegraph circuits that are reasonably well maintained. In a long circuit containing a number of intermediate repeaters in tandem, however, false operation is much more likely to occur because relatively small distortions in each of the several repeater sections may add to a high cumulative total. Under these conditions, the best answer is the employment of regenerative repeaters at the intermediate points. These are discussed in the next Article.
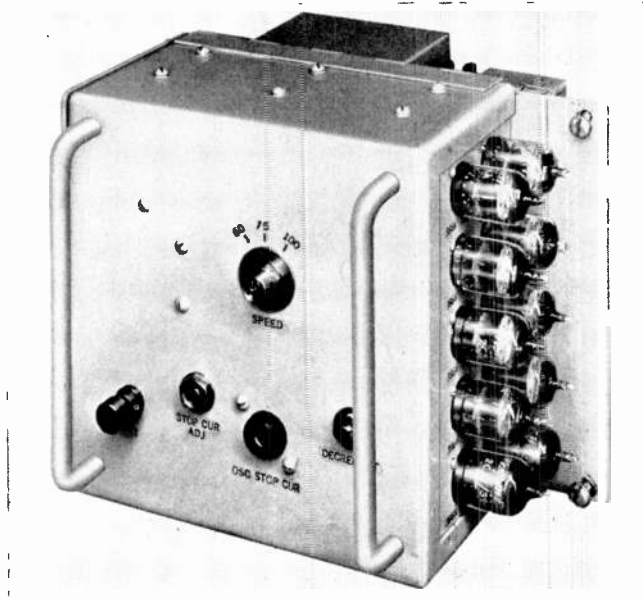
11.7 The Regenerative Repeater

The primary function of the regenerative telegraph repeater is to retime and retransmit received signal impulses restored to their original strength. It is capable of receiving without error any set of signals that would be satisfactorily received by an ordinary teletypewriter, and of sending these same signals out to the next section of line in exactly the same form as the signals produced by the sending teletypewriter at the circuit terminal.

As in the case of the teletypewriter itself, there are various designs of regenerative repeaters, including electronic as well as purely mechanical devices. The principle of operation is the same in all cases, however, and may be most readily followed by considering an early type of repeater which employed a flat distributor face with a brush arm rotating in synchronism with that of the sending teletypewriter. Such a device is illustrated schematically in simplified form by Figure 11-14. Here it will be noted that the face includes four rings of commutator segments. The two outer rings, which constitute the receiving face, may be shorted together by a pair of rotating brushes at the same time that the two inner rings, comprising the sending face, are shorted together by another pair of brushes mounted on the same rotating brush arm.

To follow the operation in detail, assume for example, that the letter *R* is to be transmitted. As indicated in Figure 11-4, the incoming signals will then consist of the starting spacing pulse, a space, a mark, a space, a mark, a space, and the final long marking pulse. When the spacing start pulse is received, the brush arm is released by the action of a relay and latch not shown in the drawing. The receiving brushes first pass over a blank segment in the outermost ring and then connect the short No. 1 receiving segment to the receiving relay armature through the second ring. This occurs at the same time that the first spacing signal of the five-pulse letter code is being received from the line, and the receiving relay is therefore operated to its spacing contact which in this case is positive. The right storing capacitor, shown at the bottom of the drawing, is accordingly charged positively. In the meantime, the brushes of the sending face have moved over segment No. 7, connecting positive battery to the sending relay through the innermost ring, and so repeating the spacing start signal to the outgoing line. As the receiving brush moves off from No. 1 segment on to a blank, the sending brush moves on to No. 1 segment of the sending face. This permits the positively charged right storing capacitor to discharge through the sending relay, thus repeating the first spacing code signal. During this operation, the receiving brush moves on

ELECTRONIC REGENERATIVE TELEGRAPH REPEATER

[99]

to segment No. 2 of the receiving face and charges the left storing capacitor negatively in accordance with the incoming marking signal. This capacitor is in turn discharged through No. 2 segment of the sending face while the right capacitor is being charged through No. 3 segment of the receiving face. The alternate operation continues until all five of the received code pulses and the final stopping mark pulse have been repeated to the outgoing line. As it completes a single revolution, the brush arm is stopped until the starting pulse for another character is received.

The basic virtue of this repeater lies in the fact that the short receiving segments pick up only the mid-portion of the incoming signals thereby allowing for considerable distortion, while the signals sent out are all of the same uniform length as those originally generated at the terminal sending teletypewriter. It is apparent, therefore, that the repeater will receive and convert to perfect signals any signals that are good enough to cause satisfactory operation of an ordinary receiving teletypewriter. In a long circuit, each regenerative repeater automatically eliminates whatever distortion may have been introduced by the preceding line section and the signals received at the terminal therefore suffer from only such distortion as may be caused by the last section, regardless of the number of sections included in the overall circuit.

CHAPTER 12

ALTERNATING CURRENTS

12.1 Source of Alternating Emf

In taking up the study of alternating-current flow, we shall follow closely the same course as was followed in the study of direct currents. The theory will precede the applications, and step by step we shall pass from the simple circuit to the network, from the network to the transmission of electric energy, and thence to our ultimate aim, which is the application of these in electrical communications systems. But along with this procedure, we shall study wherein the nature of alternating-current work differs from that of direct-current work. Perhaps the first such difference lies in the source of emf.
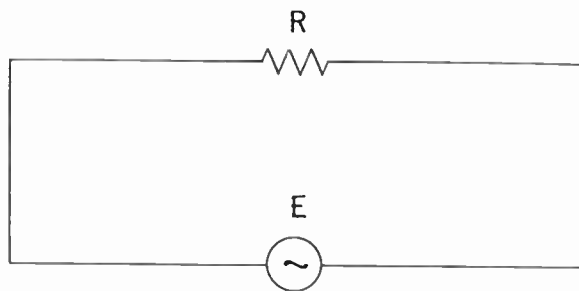

FIGURE 12-1

Figure 12-1 represents an alternating-current circuit in its simplest form. In this Figure we have a new convention for source of emf, which represents a single alternating cycle. Unlike the battery or other simple form of direct emf, we cannot completely describe such a source of emf by simply giving its voltage, for example $E = 10$ volts. Here we have a voltage which gradually increases to a maximum value, then decreases to zero, to increase again to a maximum value in the opposite direction, and then again decrease to zero, where the cycle repeats itself. Even if we knew the maximum voltage value, we should not know the trend of the successive values from zero to the maximum value. Figure 12-2 illustrates cycles of alternating emf's all very different in this respect.

Furthermore, we should not know the rapidity with which the alternations are taking place. For example, Figure 12-3 represents two cycles of identical emf values, but in one case the cycle is completed in one-half the time required for the other. Therefore to describe electrically a source of alternating emf we must know the following:

a. **The wave shape of the alternating cycle.**
b. **The value of the emf at some specified point in the cycle.**
c. **The length of time to complete the cycle, or the frequency of its repetition in a given time.**

In classifying electric currents in Chapter 8, we named two steady state conditions for alternating current; one where the wave shape is a sine wave and the other where the wave shape is not a sine wave but a complex wave. The basic study of steady-state alternating-current circuits deals with sine waves. Complex waves may be analyzed into combinations of sine waves of different frequencies, just as complex tones are analyzed into **fundamentals and harmonics.**

12.2 The Sine Wave

The sine wave is named from a trigonometric function of an angle. We have learned how it may be constructed graphically, and we may treat it as a "pattern" having a name with a mathematical origin to which an emf or current may or may not conform, rather than as a mathematical expression requiring a thorough knowledge of trigonometry for interpretation. It has interesting properties and is the natural wave form in practically all vibratory motion. It greatly simplifies alternating-current circuits because—a **sine wave emf impressed upon a circuit having a network of any number and arrangement of resistances, inductances, and capacitances with fixed values, will set up a sine wave current in every branch of the network.** No other wave shape (excepting that of direct current) will give the same wave shape for the current as that for the impressed emf.

The above rule holds in all its applications since the sine wave possesses the following properties:
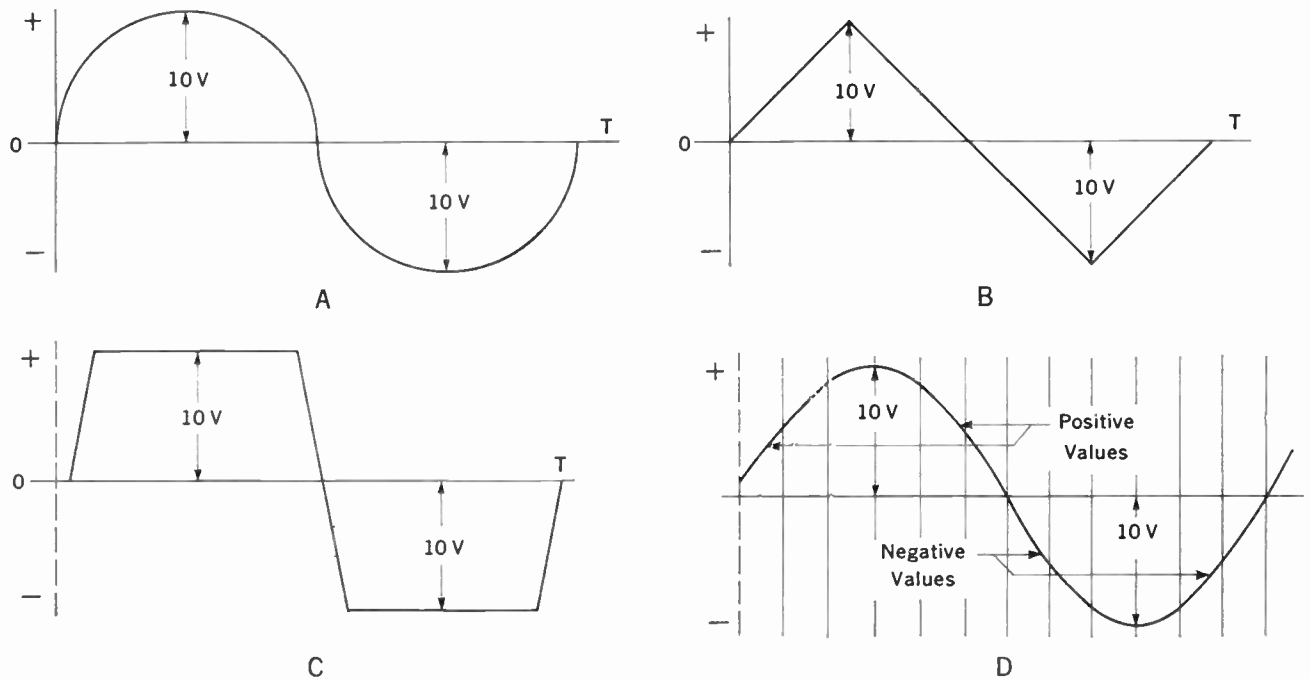
a. **Sine waves of the same frequency can be**

FIG. 12-2 THEORETICAL ALTERNATING CURRENT WAVESHAPES COMPARED WITH SINE WAVE

added (or subtracted) either in or out of "phase" and the wave shape of the result will be a sine wave. (Phase relations are defined in the next Article.)

b. A sine wave emf across a resistance, inductance or capacitance gives a sine wave current through the resistance, inductance or capacitance (though not necessarily in phase).

c. Whenever an emf is induced on account of the ever-changing value of a sine wave current, this induced emf is a sine wave (though not in phase).

12.3 Phase Relations and Vector Notation

To illustrate what is meant by phase relation, we may well discuss a method of graphically representing alternating currents and alternating emf's with vectors. Figure 12-4 shows the graph-

ical construction of a sine wave as described in Chapter 6. In this figure the horizontal scale (abscissa) represents time and the vertical scale (ordinate) represents instantaneous values of current. The complete curve then shows the values of the current for all instants during one complete cycle. It is convenient and customary to divide the time scale into units of "degrees" rather than seconds, considering one complete cycle as being completed always in 360 degrees or units of time (regardless of the actual time taken in seconds). The reason for this convention will be obvious from the method of constructing the sine curve as illustrated in Figure 12-4, where to plot the complete curve, we take points around the circumference of the circle through 360 angular degrees. It needs to be kept in mind that in the sense now used, the degree is a measure of time in terms of the frequency, and not of an angle.
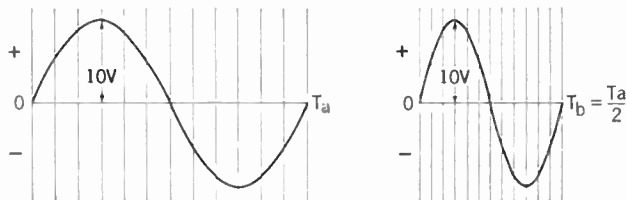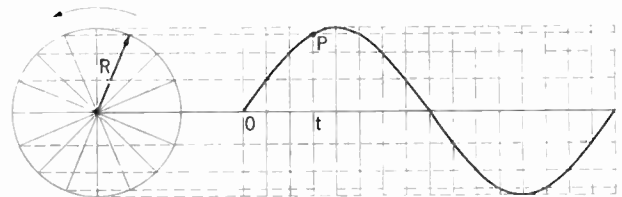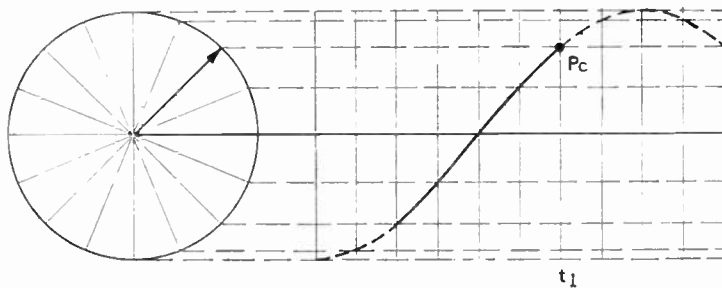

FIG. 12-3 SINE WAVES OF DIFFERENT FREQUENCY

FIG. 12-4 VECTOR REPRESENTATION OF INSTANTANEOUS CURRENT VALUE

[102]

Having adopted this convention, it is not necessary to draw the complete sine curve figure whenever we wish to represent the current in a circuit at a particular instant—for example, that current at the instant t, represented by the point P. If we know the frequency, and the length and the position of the single radius R corresponding to the point P, we have all the information we need to define the current. Here we have what we call a vector, which we can imagine as a radius of the circle, having a length equal to the maximum current or emf value of the sine wave in question. The angle this vector makes with the horizontal gives the position of point P. If we assume a direction of rotation for the vector, the instantaneous value of current at that time is measured by the vertical distance Pt which is equal to the length of the vector times the sine of the angle which it makes with the horizontal. Furthermore, we can determine by the position of the vector whether the value of the current or emf is increasing or decreasing, and its direction. The accepted convention for direction of rotation is counter-clockwise and will be understood hereafter, without the arrow being used to indicate it.
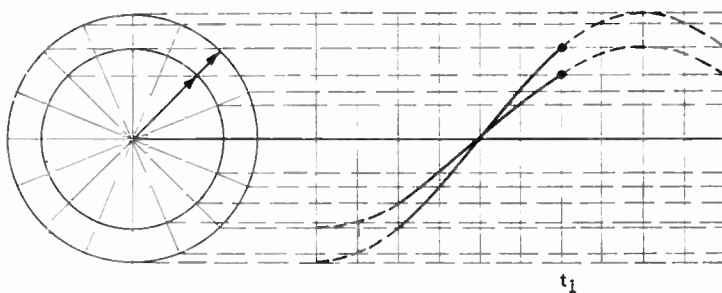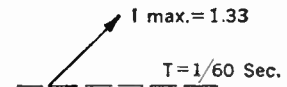
In Figure 12-1, let us assume that the maximum value of E is 10 volts, the frequency is 60 cycles per second, and the value of R is 7.5 ohms.
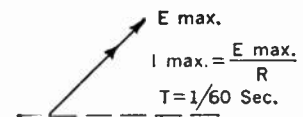

A - Voltage
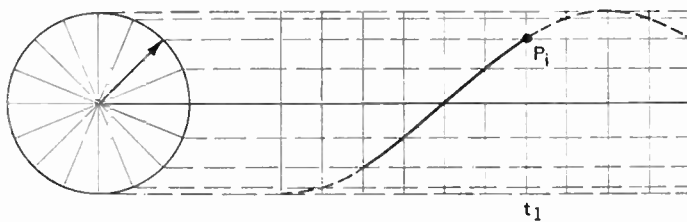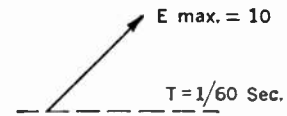
B - Current

C - Current and Voltage

E max. = 10
T = 1/60 Sec.

I max. = 1.33
T = 1/60 Sec.

E max.
I max. = $\frac{E\ max.}{R}$
T = 1/60 Sec.

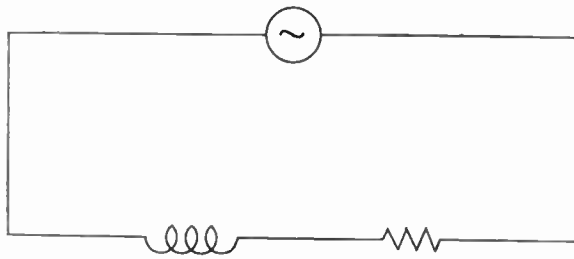VECTOR REPRESENTATION

FIG. 12-5 CURRENT AND VOLTAGE IN PHASE

[103]

Also let us assume the circuit to have negligible capacitance and inductance. By arbitrarily adopting a scale, we can represent the emf at a given instant by Figure 12-5(A). Since the inductance and capacitance of the circuit are negligible, the current at the corresponding instant will neither be retarded by inductance nor have a component part required to "charge" the circuit. It will be that determined solely by Ohm's Law. Consequently, it will change in value as the emf changes in value. In other words, it will "keep in step", becoming a maximum of 1.33 amperes at exactly the same time that the emf becomes a maximum of 10 volts, and becoming zero at exactly the same time that the emf becomes zero. **The conventional expression to describe this time relation between the voltage and the current is that the voltage and current are "in phase".**

But if, instead of a circuit such as that shown by Figure 12-1, we have the circuit shown by Figure 12-6, it will be necessary to consider the effect of the inductance. This reacts to any change in current value, and an alternating current is changing in value at all times. We should therefore expect the inductance to affect materially the value of the current and to throw the maximum points out of step, or phase, because the maximum value of current will not have been established until some time after the emf has reached its maximum value. Figure 12-7 represents the relation of voltage and current that are out of phase due to the circuit having inductance.
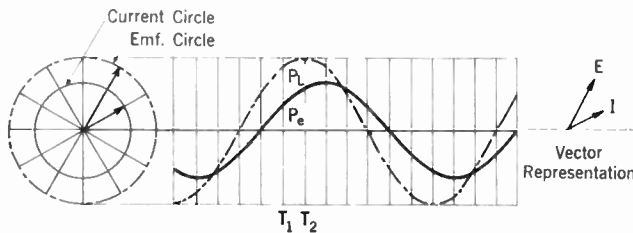

FIG. 12-7 CURRENT LAGGING IMPRESSED EMF

Here the vectorial representation must show **the extent to which the voltage and current are out of phase. This is accomplished by having the voltage vector ahead of the current vector in its rotation (in the conventional counter-clockwise direction) by an angle which is a measure of the time by which the current "lags" behind the voltage, and whose value is obvious from the relative positions of the radii of the two circles.**

In the case of a circuit having a series capacitance instead of an inductance, the circuit reactions are the reverse. The current vector then is ahead of, or "leads", the emf vector as shown by Figure 12-8. Electrical conditions in circuits containing inductance or capacitance, therefore, can be represented by current and voltage vectors, which will, in general, be out of phase.
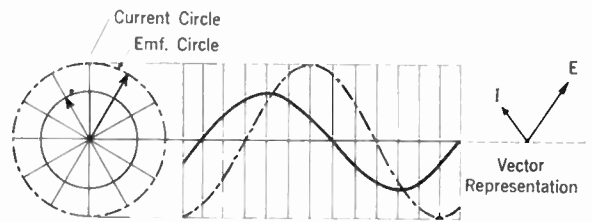
FIG. 12-8 CURRENT LEADING IMPRESSED EMF

In direct-current networks, we used equations based on Kirchoff's Laws which called for adding or subtracting current or emf values. In alternating-current work, we cannot accomplish this by merely adding the numerical lengths of the vectors. We must instead combine them in such a manner as to take into consideration any phase differences that may exist. This may be done graphically by placing the vectors to be added end to end, and drawing a line from the butt of the first arrow to the tip of the last. This line, called the resultant, is a vector which gives the magnitude and phase of the sum. For example, let us assume that it is desired to find the current delivered by the generator of Figure 12-9, when the currents in the parallel branches have the values and phase relationships indicated by vectors 1, 2, and 3. These vectors are placed end to end and the resultant drawn as indicated in 4. The length of this resultant vector gives the value of the current delivered by the generator and its angular position indicates its phase relationship with respect to the currents in the parallel branches.

12.4 Effective Emf and Current Values

In laying out current and voltage vectors thus far, we have indicated in each case the current or voltage at some particular instant of time in its cycle. The length of the vector gave the maximum value of the current or voltage and the angle that the vector made with the horizontal, in a counter-clockwise sense, indicated the particular instant being considered.

For practical purposes, however, it would be inconvenient to always state both a value and a position in time in defining an alternating current or voltage. It is advantageous, rather, to adopt some arbitrary standard so that only the value of the current or voltage need be given to define it, its position in time being understood from the convention adopted. The maximum value would perhaps appear to be the logical choice, but this has certain disadvantages. Another, and more useful value would be the average value over a complete half-cycle, this being equal for the sine wave to .636 times the maximum value.

Still more useful is a value so selected that the heating effect of a given value of alternating current in a resistor will be exactly the same as the heating effect of the same value of direct current in the same resistor. The advantage of such a convention is apparent, since it obviates to a degree the necessity for thinking of the effects of alternating and direct currents as different. This value is known as the **effective** value and is equal to the square root of the average of the squares of the instantaneous values over a complete cycle. This evaluates to .707 times the maximum value, or—

$$I = .707 \ I_{max.} \qquad (12:1)$$

and

$$E = .707 \ E_{max.} \qquad (12:2)$$

where E and I without subscripts indicate effective values. **Unless specifically stated otherwise, values of alternating currents and voltages are always given in terms of their effective values.** Likewise, in most practical applications, vectors representing currents and voltages give the effective value of the current or voltage by their length and, unlike the vectors we have previously considered, do not indicate by their angular position a particular instant of time within the cycle but only the phase relationship of the current and voltage with reference to each other, or to some other current or voltage in the same circuit.
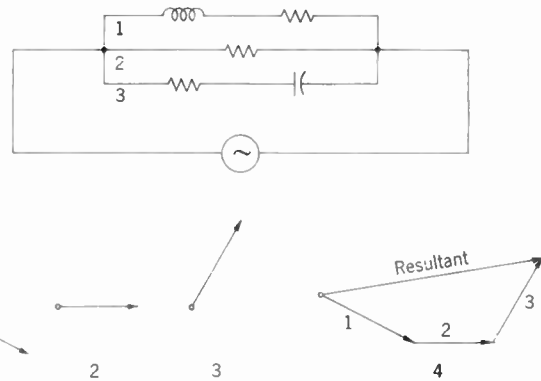
FIG. 12-9 GRAPHICAL ADDITION OF CURRENT VECTORS

12.5 Power in A-C Circuits

Just as in d-c circuits, the power in an a-c circuit is at any instant equal to the product of the current and voltage in the circuit at that instant, or we may write—

$$p = ei \qquad (12:3)$$

where the lower-case letters mean that the values are instantaneous ones. The power in an a-c circuit may, then, be shown by a curve, each point of which is obtained by taking the product of the current and voltage at the same instant of time. Such a curve for the case where the current and voltage in a circuit are in phase is shown by Figure 12-10.

It will be noted that, since the current and voltage are both negative at the same time, the power loops are both positive, which means that no power is being returned from the circuit to the generator. In other words, all of the power delivered by the generator is being absorbed in the resistance of the circuit. For this case, where the circuit contains nothing but resistance, the average power is equal to the product of the ef-
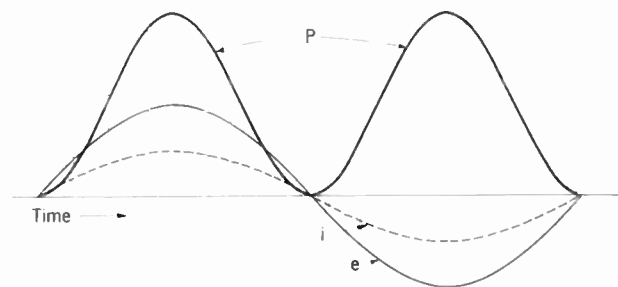
FIG. 12-10 POWER IN A-C CIRCUIT WHEN CURRENT AND VOLTAGE ARE IN PHASE

fective current and voltage, or we may write—

$$P = EI \qquad (12:4)$$

and, as always,—

$$P = I^2R \qquad (12:5)$$

The condition where the circuit contains either inductance or capacitance in addition to resistance, and the current and voltage are accordingly not in phase, is somewhat different. The power curve for such a case is shown by Figure 12-11. Here the product ei gives both positive and negative values and we have the positive power loops A and B and the negative loops C and D. The latter loops represent power returned to the generator from the circuit. The total power absorbed by the circuit is obviously equal to the sum of A and B minus the sum of C and D. In this case, then, the power, P is no longer equal to EI but to something less than that. The factor by which EI must be reduced to obtain the true power is determined by the phase relation between the current and the voltage. The power now is—

$$P = EI \cos \theta \qquad (12:6)$$

where θ is the phase angle between the current and voltage. The term, $\cos \theta$, is known as the **power factor** and has a value of 1 when θ is zero, or the current and voltage are in phase. This is the situation obtained in Figure 12-10. When θ is 90 degrees, or the voltage and current are one-quarter cycle out of phase, $\cos \theta = 0$ and power consumption is zero. For intermediate phase angles $\cos \theta$ will be between zero and one.

It may be noted that the expression, $P = I^2R$, remains true in this case and conforms with equation (12:6) because as we shall learn, $R = Z \cos \theta$ and $I = E/Z$, from whence—

$$P = I^2R = I \times I \times R$$
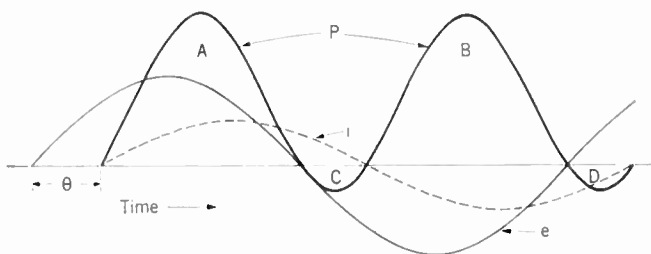$$= I \times E/Z \times Z \cos \theta = IE \cos \theta.$$

FIG. 12-11 POWER IN A-C CIRCUIT WHEN CURRENT AND VOLTAGE ARE NOT IN PHASE

12.6 Ohm's Law and Alternating-Current Calculations

In Chapter 2 we learned that the relation between the voltage and current in a d-c circuit was expressed by Ohm's Law, or

$$\frac{E \text{ (volts)}}{I \text{ (amperes)}} = R \text{ (ohms)}$$

We found this expression indispensable in our study of direct-current circuits, and certainly we shall want to apply it to alternating-current circuit calculations if we can. On the other hand, we have learned of circuit properties other than resistance that influence alternating-current flow. Moreover, these properties, viz., capacitance and inductance, not only change the value of the current in amperes but introduce changes in the phase relation of the current to the voltage. Again, the effects of inductance and capacitance depend entirely upon the particular frequency which we wish to consider. We must therefore

TABLE V

Conventional Symbols Commonly Used in Alternating-Current Work

Symbol	Stands for
P	Average power for a cycle of emf and current.
E	Effective emf.
I	Effective current.
E_{arc}	Average emf.
I_{arc}	Average current.
e	Emf at some specific instant, or instantaneous voltage.
i	Current at some specific instant, or instantaneous current.
E_1	Induced emf.
T	Length of time in seconds (or fraction of one second) for a complete cycle.
f	Frequency or the number of cycles per second.
Z	Impedance in ohms.
X_L	Inductive reactance in ohms.
X_c	Capacitive reactance in ohms.
Y	Admittance in mhos.
θ	Angle between current and impressed emf, or between impedance and resistance, etc.

introduce some new quantity that will express in ohms not only the resistance to current but the
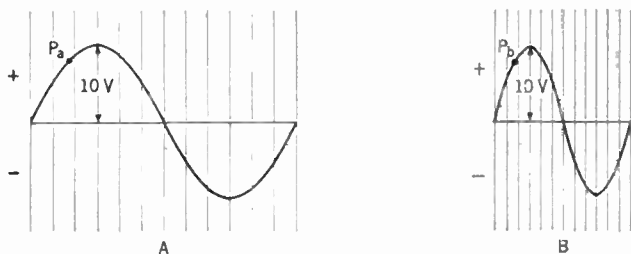
FIGURE 12-12

combined effects of resistance, capacitance and inductance at a definite stated frequency. This quantity is called impedance, and Ohm's Law is adjusted to read—

$$Z \text{ (ohms)} = \frac{E \text{ (volts)}}{I \text{ (amperes)}} \qquad (12:7)$$

where Z is the symbol for impedance or the combined effect of the circuit's resistance, inductance and capacitance taken as a single property which can be expressed in ohms for any given sine wave frequency. It follows, then, that if we can by certain calculations reduce a circuit's resistance expressed in ohms, its inductance expressed in henrys, and its capacitance expressed in microfarads, to a single expression in ohms, we can calculate the current at a given frequency in any single branch as readily as though it were a branch of a direct-current network.

The effect of inductance or capacitance in opposing the flow of current in any alternating-current circuit is known as reactance and is expressed in ohms the same as resistance. However, in combining resistance and reactance into a single property measured in ohms, which we have already referred to as impedance, we must add them vectorially because they do not act in phase. We shall take up the calculation of impedance after first learning how the reactance may be determined for any single frequency from the inductance and capacitance values in a given circuit branch.
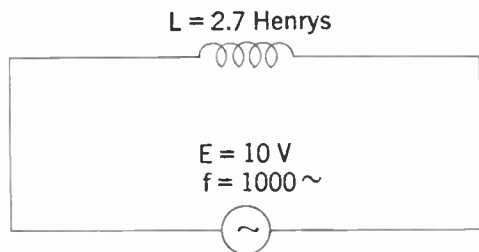
L = 2.7 Henrys

E = 10 V
f = 1000 ~

FIGURE 12-13

12.7 Inductive Reactance

Referring to Chapter 8, it will be recalled that we consider two factors as being involved in the calculation of the effects of inductance; first, the physical property of the circuit called inductance and second, the rate of change of current value, which uses inductance "as a tool" in creating the reactive effects. In an alternating-current circuit containing inductance, therefore, we should expect greater reactance for higher frequencies because higher frequencies mean an increase in the average rate of change of current. By referring to Figure 12-12 this becomes apparent. Here are two current cycles of the same effective value but the A cycle has twice the period, or half the frequency of the B cycle. Also the slope of the A curve at any point such as P_a, is half the slope at any corresponding point such as P_b, on the B curve. The slope is the measure of current change and we would expect, therefore, that the induced emf of the B curve would be twice as great as that of the A curve. Thus, the reactance due to inductance depends upon first, the inductance of the circuit and second, the frequency of the current. As a matter of fact, it can be proven that the inductive reactance expressed in ohms is equal to the inductance in henrys times the frequency in cycles per second, multiplied by 2π or—

$$X_L = 2\pi f L \qquad (12:8)$$

where X_L is the inductive reactance in ohms, π is 3.1416, f is the frequency expressed in cycles per second, and L is the inductance in henrys.

For practical use this becomes—

$$X_L = 6.2832 f L \qquad (12:9)$$

Example: In Figure 12-13 assume that the source of alternating emf is a sine wave, 10 volts, 1000 cycles per second, and the inductance shown has negligible resistance. What is the effective current through the inductance?

Note:—In practice inductance coils have appreciable resistance because any coil winding must contain a definite length of wire; the condition assumed here is that the effect of the inductance is so much greater than that of the resistance that we may neglect the value of the resistance in the calculations.
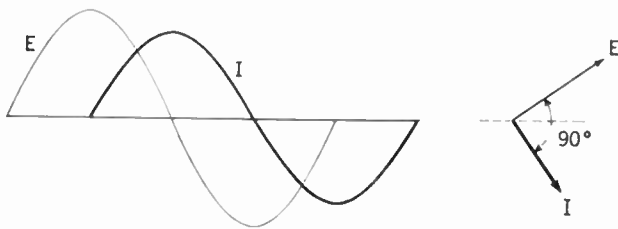
[107]

FIG. 12-14 EFFECT OF INDUCTIVE REACTANCE

Solution:

$$X_L = 6.2832fL = 6.2832 \times 1000 \times 2.7$$

$$= 16964 \text{ ohms}$$

$$I = \frac{E}{16964}$$

$$= \frac{10}{16964}$$

$$= .00059 \text{ ampere, ans.}$$

In this example, the current will be 90° behind the impressed voltage, as shown in Figure 12-14, because the induced emf due to the current must be equal and opposite to the impressed emf and the induced emf is the rate of change of current times the inductance and must, therefore, be 90° behind the current.

12.8 Capacitive Reactance

The effect of capacitive reactance in the circuit is exactly opposite to that of inductive reactance —in fact, the two tend to neutralize each other. Capacitive reactance decreases with increasing frequency and capacitance values. It also tends to make the current lead instead of lag the voltage (see Figures 12-14 and 12-16). Accordingly, if inductive reactance is assumed as positive, capacitive reactance must be taken as negative.

This time relation of the voltage and current in a circuit containing capacitance may be seen by referring to Figures 12-15 and 12-16. Here when the impressed voltage E is at its maximum positive value, the capacitor is charged to a value equal and opposite to the impressed voltage. The current in the circuit is therefore zero. As the positive impressed voltage decreases toward zero, the opposite voltage of the capacitor forces current to flow in a negative direction. This negative current reaches its maximum value when the impressed voltage becomes zero. Now the impressed voltage reverses, becoming negative, and as it

rises to its maximum negative value, it charges the capacitor in the opposite (positive) direction. During this time, the negative current decreases to zero as the capacitor becomes fully charged. Then as the negative impressed voltage decreases from its maximum, the capacitor voltage again takes control and causes the current to build up in the opposite direction. The relationships are therefore as shown in the Figure with the current leading the voltage by 90°.

The equation for capacitive reactance is as follows:

$$X_c = -\frac{1}{2\pi fC} \qquad (12:10)$$

where C is capacitance in farads. Converting C to the customary capacitance unit, microfarad, we have—

$$X_c = -\frac{1,000,000}{2\pi fC} \qquad (12:11)$$

or with 3.1416 substituted for π—

$$X_c = -\frac{1,000,000}{6.2832fC} \qquad (12:12)$$

Example: In Figure 12-15, E is 10 volts, f is 1000 and C is 2μf. What is the current in amperes?

Solution:

$$I = \frac{E}{X_c}$$

$$X_c = -\frac{1,000,000}{6.2832 \times 1000 \times 2}$$

$$= -\frac{1,000}{6.2832 \times 2}$$

$$= -79.5 \text{ ohms}$$

$$I = -\frac{10}{79.5}$$

$$= -.126 \text{ ampere, ans.}$$

(minus sign here means leading current)
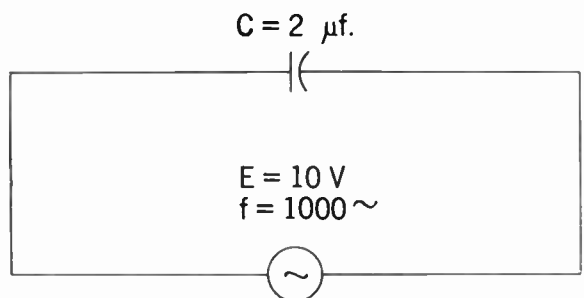

C = 2 µf.

E = 10 V
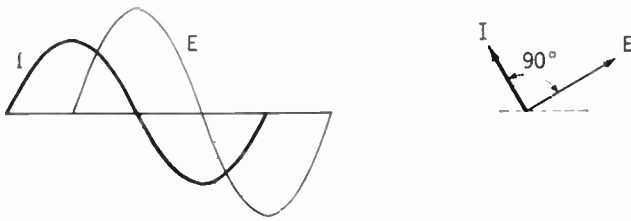f = 1000 ~

FIGURE 12-15

[108]

FIG. 12-16 EFFECT OF CAPACITIVE REACTANCE

12.9 Combination of Inductive and Capacitive Reactance

If we wish to get the combined or total reactance of an inductance in series with a capacitance, such as that shown in Figure 12-17, we may combine the reactances as follows:

$$X = X_L + X_c$$

or, from equations (12:9) and (12:12) —

$$X = 6.2832fL - \frac{1,000,000}{6.2832fC} \qquad (12:13)$$

Here the signs take care of the neutralizing effect and if the calculated value of X is positive, the inductive reactance predominates; if negative, the capacitive reactance predominates.

Example: Calculate the current in the circuit shown by Figure 12-17.

Solution:

With no resistance in the circuit—

$$I = \frac{E}{X}$$

and

$$X = X_L + X_c$$

$$= 6.2832fL - \frac{1,000,000}{6.2832fC}$$

$$X = 6.2832 \times 1000 \times .6 - \frac{1,000,000}{6.2832 \times 1000 \times 1}$$

$$= 3770 - 159$$

$$= 3611 \text{ ohms}$$

$$I = \frac{10}{3611} = .0028 \text{ ampere, ans.}$$

12.10 Impedance

To determine a way to combine reactance and resistance when we wish to evaluate the impedance, let us consider the relation between voltage

and current under two conditions; first, when a circuit contains pure resistance, and second, when it contains pure reactance. Under the first condition, we can represent the current and voltage as shown in Figure 12-18 (A), and for the second condition, as shown in Figure 12-18 (B). In a circuit containing resistance only, the current and voltage are in phase. In a circuit containing only reactance, the current either "leads" or "lags" the voltage by 90°. Thus, the effect of reactance on the phase of the current is at an angle of 90° to the effect of resistance. Therefore, in a series combination of the two, the resultant effect cannot be found by simple addition or subtraction.
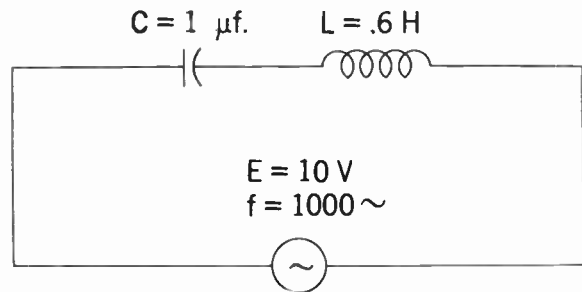
FIGURE 12-17

For the purpose of this discussion, let us assume a series combination of resistance and inductive reactance as in Figure 12-19. The vector diagram in Figure 12-20 shows a horizontal line R to represent resistance, and since the effect of reactance is at right angles to the resistance, line X_L is drawn at right angles from the end of R to represent reactance. The line Z, then represents the resultant of R and X_L, or the impedance.

Since the lines R, X_L and Z always form a right triangle, of which Z is the hypotenuse, an equation can be formed to find the value of Z directly. The equation in this case is—

FIGURE 12-18

R = 50 Ω   L = .01 H

f = 1000 ∼
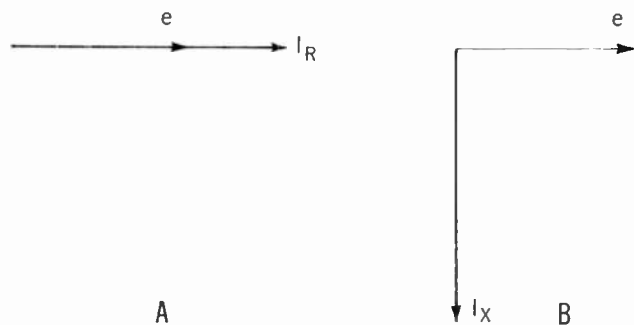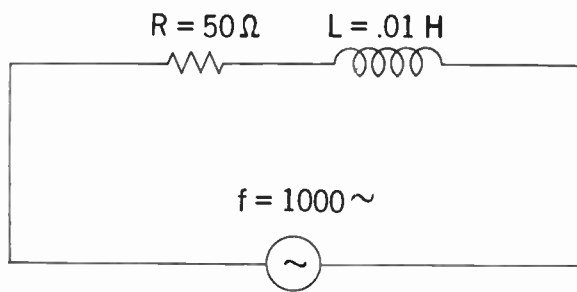
FIGURE 12-19

$$Z^2 = R^2 + X_L^2$$

or

$$Z = \sqrt{R^2 + X_L^2} \qquad (12{:}14)$$

If the reactance is a capacitive reactance, the combined effect of resistance and capacitance, as shown in Figure 12-21, may be represented by the vector diagram of Figure 12-22 in which the re-

FIGURE 12-20

actance is shown $90°$ behind the resistance since it acts in the opposite direction to inductive reactance. The mathematical result and the equation are the same as for inductive reactance—

$$Z = \sqrt{R^2 + X_c^2} \qquad (12{:}15)$$

In these diagrams if R is represented by the same line as the current, Z will be represented by the same line as the impressed emf; consequently the angle θ will represent the phase difference between the voltage and current, and with the

R = 100 Ω   C = 1 μf.

f = 1000 ∼
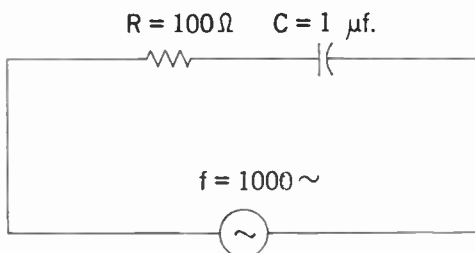
FIGURE 12-21

adopted convention for direction of rotation and that for plotting time on the sinusoidal chart, will represent current lagging behind impressed emf for positive angle as shown in Figure 12-20, and current leading impressed emf for negative angle as shown in Figure 12-22.

We can now consider a simple series circuit with all three properties, resistance, inductance, and capacitance, as shown in Figure 12-23. Here we have two reactances acting in opposite phase as shown in Figure 12-24(A). In constructing
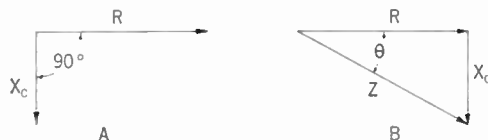
FIGURE 12-22

the impedance triangle, X_c must be considered as negative and subtracting from X_L as shown in Figure 12-24(B). If X_c is less than X_L, X will be positive, and if X_c is greater than X_L, as shown in Figure 12-25, X will be negative.

Having the relation of impedance to its component parts fixed in mind by the foregoing graphical construction, we can calculate its value in the same manner as we calculate the length of

R = 20 Ω   L = .04 H   C = 1 μf.

f = 1000 ∼
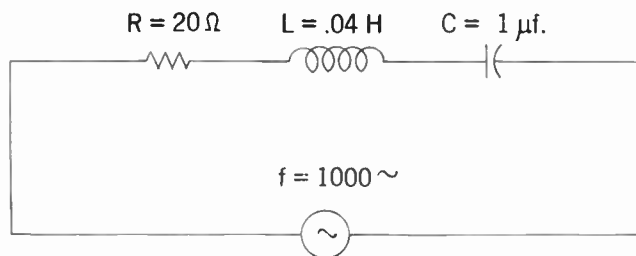
FIGURE 12-23

the hypotenuse of any right triangle, as has been explained. Thus, we may combine resistance, inductive reactance and capacitive reactance in one general equation for impedance where

$$X = X_L + X_c = 2\pi f L - \frac{1{,}000{,}000}{2\pi f C}$$

and

$$Z = \sqrt{R^2 + \left(2\pi f L - \frac{1{,}000{,}000}{2\pi f C}\right)^2} \qquad (12{:}16)$$

Example: In Figure 12-23, R is 20 ohms, f is 1000 cycles per second, L is .04 henry and C is 1µf. What is the numerical value of the impedance in ohms?

Solution:

$$Z = \sqrt{R^2 + \left(2\pi fL - \frac{1,000,000}{2\pi fC}\right)^2}$$

$$= \sqrt{(20)^2 + \left(6.28 \times 1000 \times .04 - \frac{1,000,000}{6.28 \times 1000 \times 1}\right)^2}$$

$$= \sqrt{(20)^2 + (251 - 159)^2}$$

$$= \sqrt{400 + 8464}$$

$$= \sqrt{8864}$$

$$= \quad 94 \text{ ohms, ans.}$$

TABLE VI

CHART OF VECTOR RELATIONS
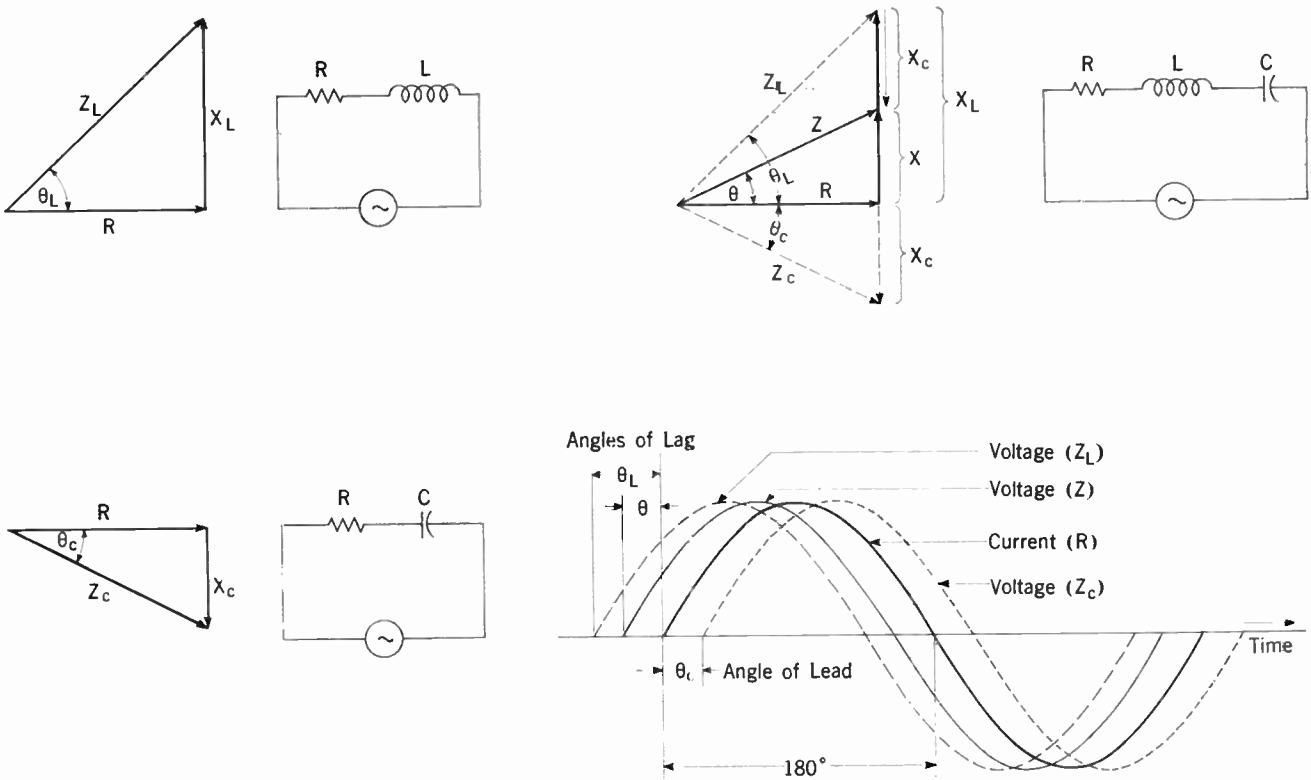
PROPERTY	REACTANCE	IMPEDANCE	PHASE ANGLE
Inductance (L)	$X_L = 2\pi fL$	$Z_L = \sqrt{R^2 + X_L^2}$	$\theta_L = \text{Tan}^{-1} \dfrac{X_L}{R}$
Capacitance (C)	$X_C = -\dfrac{1,000,000}{2\pi fC}$	$Z_C = \sqrt{R^2 + X_C^2}$	$\theta_C = \text{Tan}^{-1} \dfrac{X_C}{R}$
Net Effect	$X = X_L + X_C$	$Z = \sqrt{R^2 + X^2}$	$\theta = \text{Tan}^{-1} \dfrac{X}{R}$

NOTES: 1. If lines Z_C, Z_L or Z represent phase of voltage, line R will indicate lead or lag of current and θ_C, θ_L and θ will be angle of lead or lag.

2. Power factor is cosine of phase angle (Power $= EI \cos \theta$).

3. The impedance symbol is usually written $Z\underline{/\theta}$; for example, $Z\underline{/\theta} = 15\underline{/30°}$, etc.
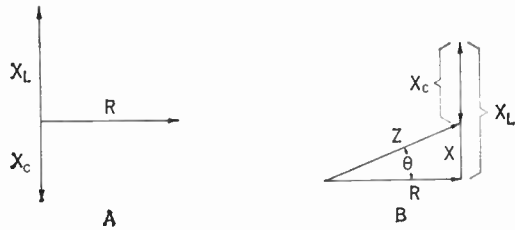
FIGURE 12-24
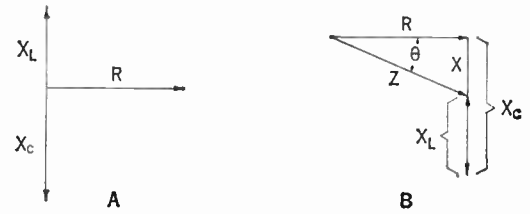
FIGURE 12-25

In these calculations we have only determined the numerical value of the impedance. This does not completely describe it, however, since there could be any number of resistance, capacitance and inductance combinations which would give the same numerical value. It is essential, therefore, to include an additional factor which will indicate the relative magnitudes of the resistance and reactance components of the impedance, in order to completely define it. This factor is the angle shown as θ in Figures 12-24 and 12-25. Impedance is customarily expressed, accordingly.

in the form $Z\underline{/\theta}$ (Z at an angle θ) where Z is the magnitude of the impedance and θ is the angle of lag or lead between any emf impressed across the impedance and the resultant current. As may be seen from Figure 12-24,

θ is equal to $\tan^{-1}\dfrac{X}{R}\left(\text{the angle whose tangent is}\dfrac{X}{R}\right)$

Also, by simple trigonometry we know that $R = Z$ $\cos\theta$ and $X = Z\sin\theta$. Thus, with the impedance expressed in the form $Z\underline{/\theta}$ it is completely defined and we may readily determine the magnitude of its resistance and reactance components.

[112]

CHAPTER 13

ALTERNATING CURRENTS—(*Continued*)

13.1 Series Networks.

In Chapters 2 and 3, means of solving direct-current networks for the current values in the various branches were described. **The same methods and equations apply to the solution of alternating-current networks.** But in this case certain additional factors enter, which, while not making the solutions any more difficult in principle, involve an increase in the amount of mathematical work required. This is due to the fact that whereas d-c quantities (current, voltage and resistance) are of only one dimension and are therefore completely described by a single number giving their magnitude, the corresponding a-c quantities are two-dimensional (i.e., vector quantities) and both their magnitudes and their time relationships with some reference point must be used in making calculations with them.
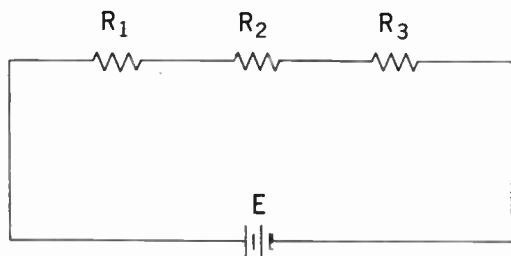
FIGURE 13-1

We learned in Chapter 3 that the total resistance in a d-c series circuit such as is shown in Figure 13-1 is equal to the **arithmetic** sum of the individual resistances, or—

$$R = R_1 + R_2 + R_3, \text{ etc.} \qquad (3:1)$$

Similarly in an a-c series circuit, as shown in Figure 13-2, the total impedance is equal to the **vector** sum of the individual impedances or—

$$\bar{Z} = \bar{Z}_1 + \bar{Z}_2 + \bar{Z}_3 \qquad (13:1)$$

the bars over the impedance symbols meaning that they are vectors and to be treated accordingly in performing the indicated additions.

To graphically illustrate the application, let us assume that $Z_1 = 10$ ohms with $\theta_1 = 30°$, $Z_2 = 15$

ohms with $\theta_2 = 45°$ and $Z_3 = 20$ ohms with $\theta_3 = 60°$; we then have the three vectors represented by Figure 13-3(A) which, when added, give the value of Z shown in Figure 13-3(B). If we should represent not only the impedance vectors but the resistance and reactance components as well, we should find that each group of components adds algebraically as shown by Figure 13-4. By comparing Figure 13-4(C) with Figure 13-4(B), we find that X is the sum of X_1, X_2 and X_3 and R is the sum of R_1, R_2 and R_3. Therefore since—

$$Z = \sqrt{R^2 + X^2}$$

we have—

$$Z = \sqrt{(R_1 + R_2 + R_3)^2 + (X_1 + X_2 + X_3)^2} \quad (13:2)$$

In order to evaluate Z, it is necessary then to find the values of the components of each individual impedance; these are obtained by multiplying the value of each impedance by the proper function of its angle.
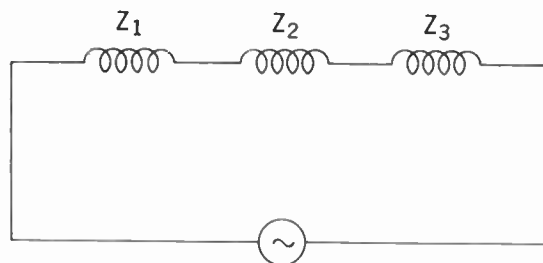
FIGURE 13-2

Example: Find the total impedance of the circuit of Figure 13-2 using the values for Z_1, Z_2 and Z_3 given above.

Solution:

$$X = Z \sin \theta$$

$$R = Z \cos \theta$$

This gives

$$X_1 = 10 \times .500 = 5 \text{ ohms}$$

$$R_1 = 10 \times .866 = 8.7 \text{ ohms}$$

[113]
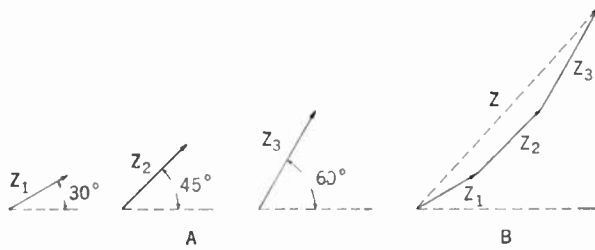
Fig. 13-3 Graphical Addition of Impedance Vectors

The other values can be determined in the same way, and we find that—

$$X_2 = 10.6 \text{ ohms}$$
$$R_2 = 10.6 \text{ ohms}$$
$$X_3 = 17.3 \text{ ohms}$$
$$R_3 = 10.0 \text{ ohms}$$

Applying equation (13:2)—

$$Z = \sqrt{(R_1 + R_2 + R_3)^2 + (X_1 + X_2 + X_3)^2}$$
$$= \sqrt{(8.7 + 10.6 + 10)^2 + (5 + 10.6 + 17.3)^2}$$
$$= \sqrt{(29.3)^2 + (32.9)^2}$$
$$= 44.0 \text{ ohms}$$

$$\theta = \tan^{-1} \frac{32.9}{29.3}$$
$$= \tan^{-1} 1.12$$
$$= 48°.$$

Therefore
$$Z = 44/48° \text{ ohms, ans.}$$

The foregoing calculation covers a general case. In practice, however, we usually have given the inductance, capacitance and resistance values rather than the individual impedances with their respective angles.

Example: Find the impedance of the series circuit shown by Figure 13-5.

Solution:

$$Z = \sqrt{(R_1 + R_2 + R_3)^2 + (X_c + X_L)^2}$$

where

$$X_c = -\frac{1,000,000}{2\pi f C}$$
$$= -\frac{1,000,000}{6.28 \times 1000 \times .2}$$
$$= -796 \text{ ohms}$$

and

$$X_L = 2\pi f L$$
$$= 6.28 \times 1000 \times .02$$
$$= 125.6 \text{ ohms}$$

Then

$$Z = \sqrt{(70 + 60 + 100)^2 + (-796 + 125.6)^2}$$
$$= \sqrt{(230)^2 + (-670.4)^2}$$
$$= 709 \text{ ohms}$$

and

$$\theta = \tan^{-1} \frac{-670.4}{230}$$
$$= \tan^{-1} (-2.9) = -71°,$$

whence

$$Z = 709/-71°, \text{ ans.}$$
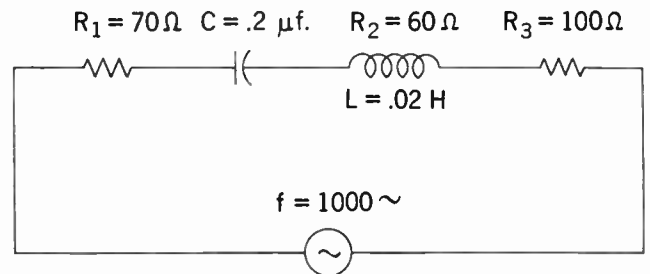
FIGURE 13-5

13.2 Parallel and Series-parallel Networks

In Chapter 3 we learned that the combined resistance of two parallel resistances was equal to—

Fig. 13-4 Analysis of Impedance Vector Addition

[114]

$$R = \frac{R_1 R_2}{R_1 + R_2} \qquad (3:4)$$

or that if more than two resistances are in parallel, the combined resistance may be found by adding together the reciprocals of each resistance (called conductance) and taking the reciprocal of this value. That is—

$$G = G_1 + G_2 + G_3 \qquad (3:6)$$

or

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

Now if we substitute impedance for resistance in the above equations, they will hold for the a-c case, providing that we remember that impedances are vector quantities. Thus for two impedances in parallel, we may write the value of the combined impedance as—

$$\overline{Z} = \frac{Z_1 \overline{Z}_2}{Z_1 + \overline{Z}_2} \qquad (13:3)$$

or, for more than two in parallel,

$$\frac{1}{Z} = \frac{1}{Z_1} + \frac{1}{Z_2} + \frac{1}{Z_3}, \text{ etc.} \qquad (13:4)$$

which latter may also be written—

$$\overline{Y} = \overline{Y}_1 + \overline{Y}_2 + \overline{Y}_3, \text{ etc.} \qquad (13:5)$$

where Y represents the reciprocal of impedance and is called **admittance.**
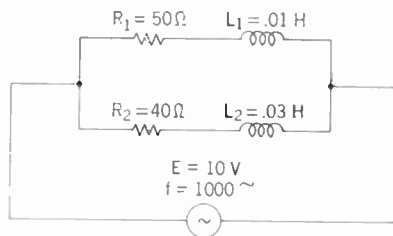
FIGURE 13-6

The mathematical solution of such equations as (13:3) ordinarily requires that the vectors be resolved into two components. These components represent resistance and reactance respectively, and are at right angles. The reactance component which is usually represented along the Y-axis is distinguished by the coefficient "j", to indicate its position relative to the resistance component along the X-axis. The vector is then expressed in the

standard notation as $\overline{Z} = Z/\underline{\theta} = R + jX$ where "j" indicates a rotation of $90°$ in a counterclockwise direction. In the algebra of complex quantities, "j" is then handled like any ordinary coefficient. The use of this notation makes possible the direct application of the same equations as those used in d-c calculations to the solution of a-c networks. As an example, let us determine the current delivered by the generator of Figure 13-6 and the phase angle of this current with the generator emf.

By Ohm's Law we know that the total current delivered by the generator is—

$$I = \frac{E}{Z/\underline{\theta}}$$

where $Z/\underline{\theta}$ is the total impedance of the circuit and consists of the net impedance of the two parallel paths whose individual impedances may be indicated as $Z_1/\underline{\theta_1}$ and $Z_2/\underline{\theta_2}$. Then from the usual equation for parallel circuits—

$$Z/\underline{\theta} = \frac{Z_1/\underline{\theta_1} \times Z_2/\underline{\theta_2}}{Z_1/\underline{\theta_1} + Z_2/\underline{\theta_2}} \qquad (13:3)$$

The first step is to find the values of $Z_1/\underline{\theta_1}$ and $Z_2/\underline{\theta_2}$. We know that—

$$\theta_1 = \tan^{-1} \frac{X_1}{R_1}$$

where

$$R_1 = 50 \text{ ohms}$$

and

$$X_1 = 2\pi f L_1 = 6.28 \times 1000 \times .01 = 62.8 \text{ ohms}$$

Then

$$\theta_1 = \tan^{-1} \frac{62.8}{50} = \tan^{-1} 1.255 = 51°27'$$

Then, since $R_1 = Z_1 \cos \theta_1$—

$$Z_1 = \frac{R_1}{\cos \theta_1} = \frac{50}{\cos 51°27'} = \frac{50}{.6232} = 80.3 \text{ ohms}$$

Impedance of branch 1

$$= Z_1/\underline{\theta_1} = 50 + j\,62.8 = 80.3/\underline{51°27'}.$$

Likewise $R_2 = 40$ ohms

$$X_2 = 2\pi f L_2 = 6.28 \times 1000 \times .03 = 188.4 \text{ ohms}$$

$$\theta_2 = \tan^{-1} \frac{188.4}{40} = \tan^{-1} 4.71 = 78°1'$$

$$Z_2 = \frac{40}{\cos 78°1'} = \frac{40}{.2076} = 193.0 \text{ ohms}$$

Impedance of branch 2

$$= Z_2/\theta_2 = 40 + j\,188.4 = 193.0/\underline{78°1'}.$$

Then using equation (13:3) and expressing the vectors of the numerator in the Z/θ form, since multiplication is involved, and the vectors of the denominator in the $R + jX$ form, since addition is involved, we have—

$$Z/\underline{\theta} = \frac{Z_1/\theta_1 \times Z_2/\theta_2}{(R_1 + jX_1) + (R_2 + jX_2)}$$

$$= \frac{Z_1 Z_2/\theta_1 + \theta_2}{(R_1 + R_2) + j(X_1 + X_2)}$$

$$= \frac{80.3 \times 193.0/\underline{51°27' + 78°1'}}{(50 + 40) + j\,(62.8 + 188.4)}$$

$$= \frac{15500/\underline{129°28'}}{90 + j251.2}$$

$$= \frac{15500/\underline{129°28'}}{\dfrac{90}{\cos\left[\tan^{-1}\dfrac{251.2}{90}\right]}\Big/\tan^{-1}\dfrac{251.2}{90}}$$

$$= \frac{15500/\underline{129°28'}}{\dfrac{90}{\cos(\tan^{-1}2.79)}\Big/\tan^{-1}2.79}$$

$$= \frac{15500/\underline{129°28'}}{\dfrac{90}{\cos 70°17'}\Big/\underline{70°17'}}$$

$$= \frac{15500/\underline{129°28'}}{\dfrac{90}{.3374}\Big/\underline{70°17'}}$$

$$= \frac{15500/\underline{129°28'}}{267/\underline{70°17'}}$$

$$= \frac{15500}{267}\Big/\underline{129°28' - 70°17'}$$

$$= 58.0/\underline{59°11'}$$

And

$$I = \frac{E}{Z/\theta} = \frac{10}{58/\underline{59°11'}}$$

$$= \frac{10}{58}\,/\underline{0° - 59°11'} = .173/\underline{-59°11'}$$

Thus we find that the generator will deliver a current of .173 ampere and this current will lag the generator voltage by 59° 11'.

With a little practice it will be found that several of the detailed steps given above can be per-

[116]
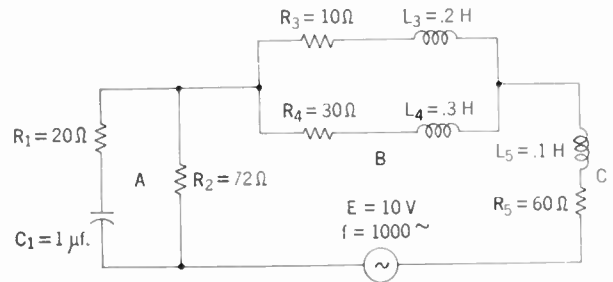
FIGURE 13-7

formed in a single operation. This may be illustrated by solving the circuit of Figure 13-7 to find the current delivered by the generator and its phase relationship with the emf.

Solution:

$$Z_1 = R_1 + jX_1$$

$$X_1 = -\frac{10^6}{2\pi f C_1} = -\frac{1,000,000}{6.28 \times 1000 \times 1} = -159.3$$

$$Z_1 = 20 - j\,159.3 = 160.7/\underline{-82°51'}$$

$$Z_2 = R_2 + jX_2 = 72 + j0 = 72/\underline{0°}$$

$$Z_A = \frac{Z_1 Z_2}{Z_1 + Z_2} = \frac{160.7/\underline{-82°51'} \times 72/\underline{0°}}{20 - j\,159.3 + 72 + j0}$$

$$= \frac{11,570/\underline{-82°51'}}{92 - j\,159.3} = \frac{11,570/\underline{-82°51'}}{184/\underline{-60°1'}}$$

$$= 62.8/\underline{-22°50'} = 58.0 - j24.4$$

$$Z_3 = R_3 + jX_3$$

$$X_3 = 2\pi f L_3 = 6.28 \times 1000 \times .2 = 1256$$

$$Z_3 = 10 + j\,1256 = 1256/\underline{89°33'}$$

$$Z_4 = R_4 + jX_4$$

$$X_4 = 2\pi f L_4 = 6.28 \times 1000 \times .3 = 1884$$

$$Z_4 = 30 + j\,1884 = 1884/\underline{89°5'}$$

$$Z_B = \frac{Z_3 Z_4}{Z_3 + Z_4} = \frac{1256/\underline{89°33'} \times 1884/\underline{89°5'}}{10 + j\,1256 + 30 + j\,1884}$$

$$= \frac{2,365,000/\underline{178°38'}}{40 + j\,3140}$$

$$= \frac{2,365,000/\underline{178°38'}}{3140/\underline{89°16'}}$$

$$= 753/\underline{89°22'} = 8 + j\,753$$

$$Z_C = R_5 + jX_5$$

$$X_5 = 2\pi f L_5 = 6.28 \times 1000 \times .1 = 628$$

$$Z_C = 60 + j\,628$$

The total impedance $Z = Z_A + Z_B + Z_C$

$$Z = 58.0 - j\,24.4 + 8 + j\,753 + 60 + j\,628$$

$$= 126 + j\,1356 = 1360\underline{/84°40'}$$

$$I = \frac{E}{Z} = \frac{10}{1360\underline{/84°40'}} = .00735\underline{/-84°40'}$$

The current delivered by the generator has a value of .00735 ampere and lags the impressed voltage by 84°40'.

13.3 Alternating-Current Resistance

In alternating-current networks, the apparent resistance of a particular piece of apparatus is often quite different from its direct current or true resistance. The resistance offered to alternating current may be much greater than that offered to direct current; furthermore, in such cases the value of the resistance depends on the alternating-current frequency. We find, then, that **not only the reactance component of an impedance but its resistance component as well may be a function of the frequency.**

"Alternating-current resistance", in the case of a coil winding for example, represents not only the actual resistance of the conductor used to wind the coil but includes also a factor due to power losses within the iron core. These losses are caused in part by hysteresis and in part by small currents induced in the iron itself as a conductor, and called **eddy currents**. The total power loss in the coil includes not only the heat losses due to the resistance of the coil winding but also the core losses. Since any power loss can be expressed in the form of the equation $P = I^2R$, we assume that the winding has in effect a resistance which satisfies this equation. The part of the power loss that is due to the iron core increases with the frequency. Therefore, we should expect the a-c resistance for a high frequency to be greater than the a-c resistance for a low frequency. Certain comparable losses may occur in the dielectric materials of capacitors that may also have the effect of increasing the apparent resistance of the circuit.

The effective resistance of a long wire such as an ordinary transmission line also increases as the frequency of the current through it increases, due to a cause known as **skin effect**. It is a result of the fact that any current flowing through a wire is accompanied by a magnetic field set up around the wire. This field exists to some degree at least

within the wire as well as in the space surrounding it. As the frequency of the current in the wire increases, however, the reaction in the interior of the wire between the current flowing and the counter-voltages induced by the magnetic field tends to prevent the field from penetrating deeply into the wire. The net effect is to force the current in the wire to flow nearer the surface instead of being evenly distributed through the cross-sectional area of the conductor. Since the current thus flows through an effectively smaller conductor area the resistance increases. At low frequencies this effect is small but at the high frequencies used in radio work the a-c resistance of a conductor due to skin effect may be many times the d-c value.

13.4 Resonance

In a circuit containing a given inductance, the reactance, X_L, depends upon the frequency; if the frequency is doubled, the reactance is also doubled. In the case of a given capacitance value,
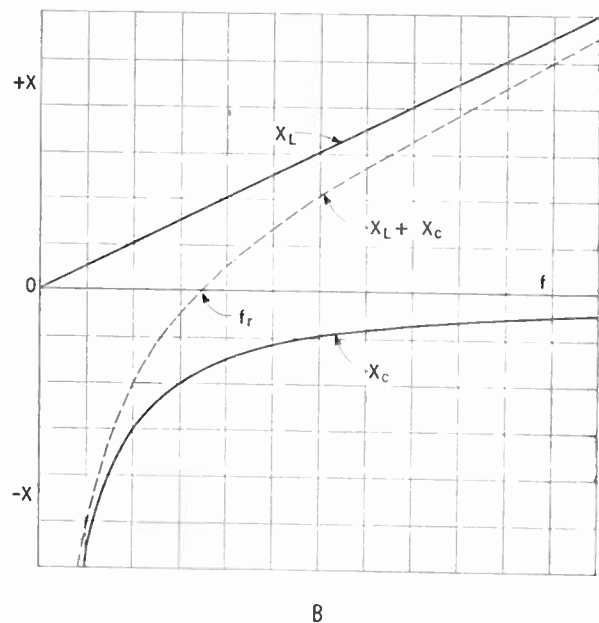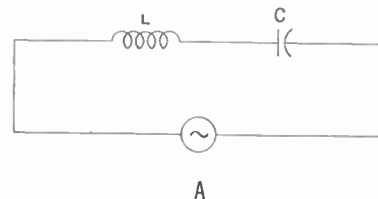
A
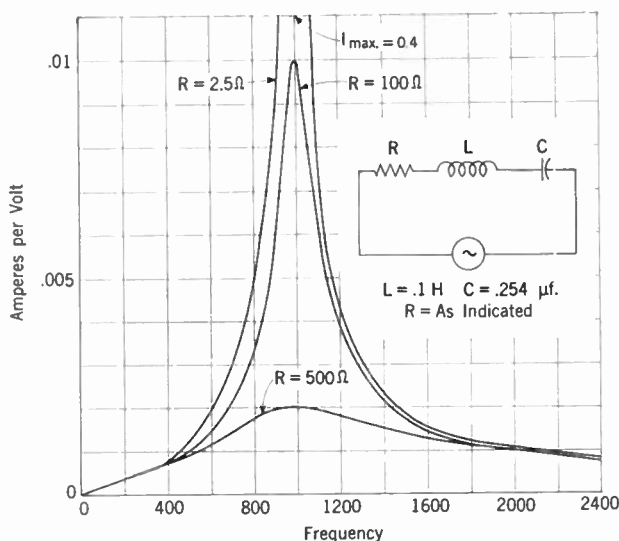
B

FIG. 13-8 SERIES RESONANT CIRCUIT

FIG. 13-9 CURVES OF CURRENT VALUES IN SERIES
RESONANT CIRCUIT

on the other hand, the negative reactance, X_c, is reduced when the frequency is increased. This is illustrated by Figure 13-8 where the inductive reactance X_L and the capacitive reactance X_c are plotted against frequency. If a series circuit contains both inductance and capacitance, as shown in Figure 13-8 (A), there is therefore some frequency at which the negative reactance, X_c, becomes equal but opposite in value to X_L. Where the dotted line crosses the zero axis in Figure 13-8 (B), the combined reactance is equal to zero. **This is called the frequency to which the circuit is resonant, or where—**

$$0 = 2\pi fL - \frac{1,000,000}{2\pi fC}$$

The value of the resonant frequency, f_r, can be determined in terms of the inductance and capacitance by solving this equation for f as follows:

$$f_r = \frac{1,000}{2\pi\sqrt{LC}} \tag{13:5}$$

Since the total reactance is zero at the resonant frequency, the impedance is then equal to the resistance of the circuit and the current flow is determined solely by this resistance value.

Figure 13-9 illustrates the behavior of a series resonant circuit similar to that shown in Figure 13-8 (A), but including some resistance, when the applied emf is varied through a band of frequencies. The curves were plotted by assuming an impressed emf of 1 volt for each frequency of the band, and three different values of resistance. As

will be noted, the peak current values depend entirely upon the resistance values, for at the peak the positive and negative reactances exactly neutralize each other and the current is determined solely by the resistance. The addition of resistance to the series resonant circuit reduces the selectivity or sharpness of the resonance peak. That is, the ratio of the current at the resonant frequency to the current at frequencies near the resonant frequency is reduced. Below the resonant frequency, the capacitance in the circuit will have the major effect in limiting the current, and the circuit will tend to look like capacitance with the current leading the voltage. Above resonance the inductance will limit the current and the circuit will look like inductance with the current lagging the voltage. At resonance the current and voltage are in phase although the current may be relatively very large. The actual voltages across the inductor and capacitor may therefore be many times as great as the voltage applied to the circuit.

Example: To what frequency is the circuit shown by Figure 13-9 resonant if C is .254 μf, L is .10 H; what current will flow at resonance when R is 4 ohms and E is 1.0 volt and what is the voltage drop across the inductance?

Solution:

$$f_r = \frac{1,000}{6.28\sqrt{.10 \times .254}}$$

$$= \frac{1,000}{6.28\sqrt{.0254}}$$

$$= \frac{1,000}{6.28 \times .159}$$

$$= 1,000 \text{ cycles per sec. Ans.}$$

$$I = \frac{E}{R} = \frac{1.0}{4} = .25 \text{ amp. Ans.}$$

$$E_L = IX_L$$

$$X_L = 2\pi fL = 6.28 \times 1,000 \times .10 = 628$$

$$E_L = .25 \times 628 = 157 \text{ volts. Ans.}$$

The resonance principle has numerous and interesting uses in connection with communication circuits. One application is the use of a capacitor of proper value in series with a telephone receiver winding, repeating coil winding, or other winding having inductance, where it is desired to increase the current.

[118]

A much more common use of the resonance principle is the so-called "tuned" circuit which is so extensively employed in radio and other high frequency applications. It is an arrangement whereby the circuit has a much lower impedance to some particular frequency than to any other frequency; if a band of frequencies is impressed, it selects, so to speak, a high current for the particular frequency but permits only a small current for any other frequency. Figure 13-9 illustrates this principle.

In many tuned circuits, capacitance and inductance are connected to form a **parallel-resonant** circuit as shown in Figure 13-10 (A). For this connection, when the positive reactance is equal and opposite to the negative reactance, and the resistance of the inductor is low, the combined impedance presented to the generator is extremely great and there is a minimum load on the generator. In other words, the generator circuit is practically open. Figure 13-10 (B) shows the com-

FIG. 13-11 CURVES OF CURRENT VALUES IN PARALLEL-RESONANT CIRCUIT

bined reactance, X_t, presented to the generator by this circuit. It can be seen that at the resonant frequency the two parallel reactances combine to give an extremely high value. At the same time, there must be a current through the inductance, determined by dividing the voltage of the generator by the impedance of this branch. Similarly, there must be a current through the capacitor which can be determined in the same way. These currents are equal in value, but are flowing in opposite directions, thereby neutralizing each other in the lead to the generator. Effectively, this gives an open circuit in so far as the generator is concerned, but a circuit equal to either the inductance or capacitance alone connected to the generator in so far as either of the branches is concerned. The physical explanation here is that a current is oscillating around through the inductor and capacitor, with the emf of the generator merely sustaining this oscillation. Of course, since the inductance must have some resistance, there will be an I^2R loss in the inductance, and it would never be possible to have the theoretical case where the generator current is actually zero.

Figure 13-11 illustrates the selectivity of a parallel-resonant circuit made up of the same units as were used in the series resonant circuit. It will be noted that the selectivity of the parallel-

FIG. 13-10 PARALLEL-RESONANT CIRCUIT

[119]

resonant circuit is also decreased as the resistance is increased. Indeed, there is a value of resistance beyond which the circuit loses its resonant characteristics altogether. Moreover, in this case, the resistance may be seen to have some effect on the value of the resonant frequency.

In radio and other high frequency work the parallel-resonant circuit is often called a "tank circuit". This is because it acts as a storage reservoir for electric energy. Here it may be more helpful to think in terms of energy transferring back and forth between the electric field of the capacitor and the magnetic field of the inductor rather than merely of current oscillating back and forth in the parallel circuit.

13.5 Q

Where resonant circuits are used for purposes of tuning or frequency selection, it is important that the resistance of the circuit be held to a minimum. This is illustrated by Figures 13-9 and 13-11 where it is shown that the sharpness of tuning is greater the lower the resistance. Since the resistance in the circuit is largely contained in the inductor, the objective is to have the ratio of the reactance of the inductor to its a-c resistance as high as possible. This ratio is known as the **Quality** or Q of an inductor and is usually expressed by the equation—

$$Q = \frac{X_L}{R} = \frac{2\pi f L}{R} \qquad (13:6)$$

The Q of resonant circuits used in practice varies from values in the order of a hundred in the case of iron-core coils to values as high as 20 thousand or more in the case of certain cavity resonators.

In a series resonant circuit at the resonant frequency, the voltage developed across the capacitor is Q times the net voltage across the coil and capacitor, or $E_c = QE$. In a parallel-resonant circuit when Q is large, the total impedance of the combination at the resonant frequency is Q times the coil or capacitor reactance, or $Z = X_L Q$ (at resonance $X_c = X_l$).

> **Example:** In a parallel-resonant circuit having an inductance of 50 microhenries and a resistance of 10 ohms, what is the Q of the circuit and what is the impedance at a resonant frequency of 1000 kc?

Solution:

50 microhenries $= 5 \times 10^{-5}$ henries; 1000 kc $= 10^6$ cycles.

$X_L = 2\pi f L = 6.28 \times 10^6 \times 5 \times 10^{-5} = 314$ ohms

$$Q = \frac{X_L}{R} = \frac{314}{10} = 31.4$$

$Z = X_L Q = 314 \times 31.4 = 9860$ ohms. Ans.

CHAPTER 11

REPEATING COILS AND TRANSFORMERS

14.1 Mutual Induction

The inductive effects discussed in Chapter 8 dealt with the magnetic interlinkages from one turn of a coil winding to the other turns of the same winding. We defined the effects coming from such magnetic interlinkages as **self-inductance**. The current resulting from the **induced** emf was superposed upon the current resulting from the **impressed** emf.

In practice, we may experience inductive effects in circuits other than the one in which the current due to the impressed emf is flowing. **That is to say, two coils may be so related that the lines of magnetic induction established by a current in the first coil may cut the turns of the second coil (which may be connected to an entirely different circuit) in the same way that similar lines established by any one turn of a single coil cut the other turns of the same coil. This effect is called mutual induction** and the property of the electric circuit that is responsible for the effect is known as its **mutual inductance.**
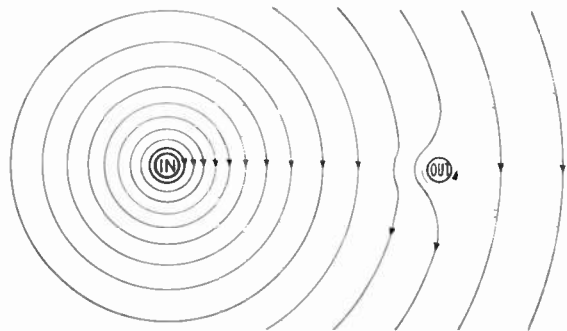
FIGURE 14-1

14.2 Theory of the Transformer

In the study of magnetism we found that a wire in which there is a current is always surrounded by a magnetic field. This field, when created by a current establishing itself in the conductor, grows outwardly from the wire as the current increases. Figure 14-1 shows a group of lines of magnetic

induction around a conductor (shown in cross-section) in which the current is increasing in value. If a second conductor is in the vicinity, it will be cut by these lines moving outward from the current-carrying conductor. This induces an emf in the second conductor, which, as illustrated in the Figure, will establish a current in the opposite direction to that in the first conductor. The induced current will cease to flow, however, when the current in the first conductor reaches its maximum value, or at any other instant when it may have a steady, unchanging value because the magnetic field has become stationary and the lines of magnetic induction move neither outward nor inward for a steady current value.
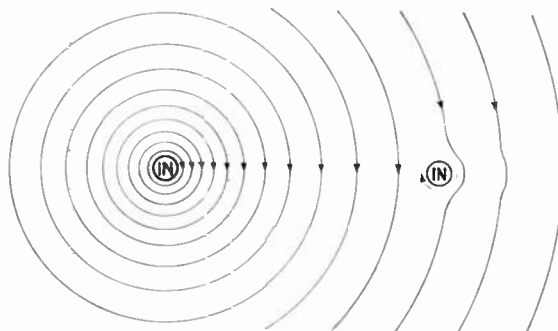
FIGURE 14-2

If the current in the first conductor is decreased, we have the reverse condition, or that shown in Figure 14-2. Here the lines, instead of expanding and moving outward, are contracting and moving inward, again cutting the second conductor, but now the current induced is in the opposite direction. It is now in the same direction in the second conductor as in the first. This law for induced emf may be expressed as follows: **For any two parallel conductors, a current in one increasing in value induces an emf in the other, tending to establish a current in the opposite direction, and a current decreasing in one will induce an emf in the other, tending to produce a current in the same direction.**

[121]
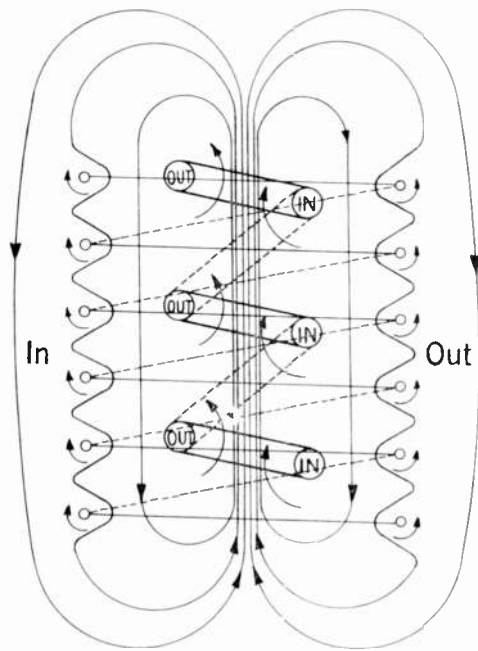
FIG. 14-3 PRINCIPLE OF INDUCTION COIL

Instead of the two single conductors shown in Figures 14-1 and 14-2, let us consider two separate coils, one inside the other, as in Figure 14-3. If we call the one carrying the original current the **primary**, which in this case we may represent by the inside coil, and the other the **secondary**, we shall find that a relatively strong magnetic field is established by a changing current in the primary. This will cut the entire group of conductors represented by the turns of the secondary, thereby inducing appreciable potential in the secondary. The ordinary telephone induction coil operates in this manner. The primary, when connected in series with the transmitter, carries a current which decreases and increases in value in response to the varying resistance of the transmitter. Consequently, an alternating current is induced in the secondary of the coil.

If now the two separate coils of Figure 14-3 are wound on the same iron core in the manner indicated by Figure 14-4, the effect will be intensified. Because the iron offers a path of low reluctance to the magnetic flux, the total number of lines will be greatly increased and all of the lines set up by the primary winding, P, will cut all of the secondary winding, S.

If the windings, P and S have the same number of turns, and both the coils and core are constructed so as to have negligible energy losses,

we shall find that the voltmeter reading is the same when connected across the terminals of S as when connected across the terminals of P. In other words, the induced emf of the secondary winding is equal to the impressed emf of the primary winding. Such a device is called an ideal transformer of **unity** ratio.

If, now, we should increase the number of turns of the secondary winding S, we would find that the voltmeter reading would be greater on the secondary than on the primary side of the transformer. If we should decrease the number of turns of the winding S, the effect would be reversed. We have here a means, therefore, of controlling the voltage applied to a load; we may effectively increase or decrease the generator voltage by a proper choice of transformer. If a transformer has a greater number of turns on the secondary than on the primary so that the voltage is increased, it is called a **step-up** transformer; if it has a lesser number of turns on the secondary than on the primary so that the voltage is decreased, it is called a **step-down** transformer. The voltage across the two windings is directly proportional to the number of turns. This relation is expressed by the equation:

$$\frac{E_P}{E_S} = \frac{N_P}{N_S} \qquad (14:1)$$

We may explain this relation between the number of turns and voltage by our original law governing inductive effects, which states that **the induced voltage is proportional to the rate of cutting lines of magnetic induction.** Each time the alternating emf in the primary completes a cycle, it establishes a magnetic flux in the iron core which collapses to be established in the opposite direction, to again collapse, etc. This flux will cut each and every turn about the iron core. In doing so, for the ideal case where there is no loss due to magnetic leakage, etc., the same voltage is induced in each individual turn. This voltage may be represented by the symbol v. Now, the voltage measured across the secondary (with no load con-
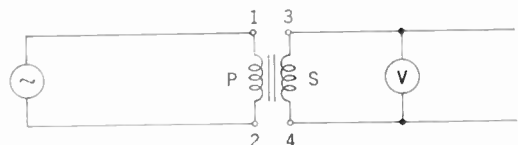

FIG. 14-4 ELEMENTARY TRANSFORMER CIRCUIT

[122]

nected) is merely the sum of these individual turn voltages or—

$$E_s = N_s \times v \qquad (14:2)$$

where N_s is the number of turns on the secondary.

In the primary the induced emf must be exactly equal and opposite to the impressed emf since the emf due to IR drop is practically negligible. This could be expressed by an equation similar to equation (14:2), thus—

$$E_P = N_P \times v \qquad (14:3)$$

Since v is the same in both equations (14:2) and (14:3), we may derive equation (14:1) by dividing (14:3) by (14:2).
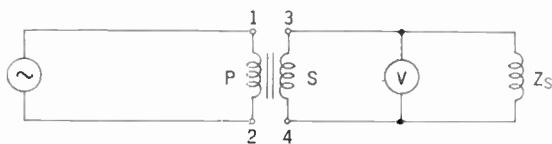

FIGURE 14-5

If a load in the form of a shunting impedance is connected to the secondary of the transformer, as shown in Figure 14-5, the induced emf in the winding S causes a current to flow through the impedance Z_s, which can be expressed as follows—

$$I_s = \frac{E_s}{Z_s}$$

When this current starts to flow through the load Z_s, and through the winding S, it will establish other lines of magnetic induction in the transformer core, which oppose those established by the current in the winding P. This will tend to neutralize the magnetic field in the iron core, thereby tending to counteract the inductance of the winding P and to make it more nearly like a plain resistance. With the induced emf in the winding P reduced, a greater current will flow from the generator through this winding, thus again increasing the flux in the iron core, so that finally there are produced the same induced emf effects as in the case of the transformer on open circuit. We therefore find that the transformer adjusts itself to any load that may be connected to the secondary just as if an equivalent load were connected directly to the generator, i.e., the current supplied by the generator increases with an increase of current in the secondary of the transformer.

The current, however, is not necessarily of the same value in the primary as in the secondary, but like the voltage, depends upon the ratio of the number of turns of the primary to the number of turns of the secondary. The relation between current values is the inverse ratio of the number of turns. In other words, the winding having the greater number of turns has a proportionately smaller current. This is seen when we consider that the flux in the core depends upon the current value times the number of turns, and the flux established by one coil balances that established by the other—

$$N_P \times I_P = N_s \times I_s \text{ or } \frac{I_s}{I_P} = \frac{N_P}{N_s} \qquad (14:4)$$

The same relation can be determined in another way. We know from the law of conservation of energy that the energy existing in the secondary circuit can never exceed, but for an ideal transformer will be just equal to, the energy of the primary circuit, where since—

$$P_P = P_s \text{ and } P = EI,$$

we have—

$$E_s I_s = E_P I_P$$

from which—

$$\frac{I_s}{I_P} = \frac{E_P}{E_s} \text{ or } \frac{I_s}{I_P} = \frac{N_P}{N_s} \qquad (14:4)$$

The value of the current in the secondary circuit is of course dependent on the value of the load impedance Z_s—that is, $I_s = E_s/Z_s$. From which—

$$Z_s = \frac{E_s}{I_s}$$

Similarly the impedance presented to the generator by the primary of the transformer is—

$$Z_P = \frac{E_P}{I_P}$$

The relationship between the impedance ratio and the turn ratio is then determined, with the help of equations (14:1) and (14:4), as—

$$\frac{Z_s}{Z_P} = \frac{E_s}{I_s} \times \frac{I_P}{E_P} = \frac{E_s}{E_P} \times \frac{I_P}{I_s} = \frac{N_s}{N_P} \times \frac{N_s}{N_P}$$

or

$$\frac{Z_s}{Z_P} = \left[\frac{N_s}{N_P} \right]^2 \qquad (14:5)$$

Inequality ratio transformers may be rated either according to their voltage ratios, step-up
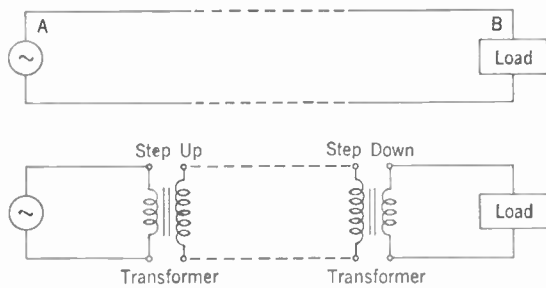
FIG. 14-6 USE OF TRANSFORMERS IN POWER CIRCUIT

or step-down as the case may be, or in accordance with their impedance ratios. In power work where transformers are primarily used to change the voltage of the system, the rating is on the voltage basis. In telephone work where inequality transformers are used in most cases primarily to match unequal impedances, as will be explained later, they are usually rated in accordance with their impedance ratios.

Before taking up specific uses of the transformer, let us review in general what its presence in Figure 14-5 has or may have accomplished.

a. The characteristics of the electric energy may have been changed, or we might say its state may have been "transformed", inasmuch as in the primary circuit we may have had high current and low voltage, while in the secondary circuit we may have had low current and high voltage, or vice versa, depending upon whether the transformer was step-up or step-down.

b. The electric energy was transferred from one circuit to another without any metallic connection being made between the two circuits; from a direct-current aspect the circuits are separate units.

c. The transformer in effect changed the nature of the connected load, or in other words changed the impedance of the load to a different value unless the transformer was of unity ratio.

In power work the principal use of a transformer is to accomplish the result given in a above, whereas in telephone work we are more often concerned with b and c.

First, let us illustrate the power case by referring to Figure 14-6 which shows the use of transformers in a simple power transmission circuit. Let us assume that a 110-volt alternating-current generator at station A is to be used to supply a load several miles away. The load is of such nature that it must have 100 amperes at an impressed voltage of 100 volts. Transmission from A to B must, therefore, be accomplished with a loss of 10 volts for a current of 100 amperes and this means that the IR drop of the line must not exceed 10 volts. Therefore, the resistance of the line, from the equation—

$$R = \frac{E}{I} = \frac{10}{100} = \frac{1}{10} \text{ ohm}$$

must not exceed 1/10th of an ohm, requiring extremely large copper conductors. If, however, a step-up transformer of 1-to-20 voltage ratio is inserted at the generator, and a step-down transformer of 20-to-1 ratio is inserted at the load, we shall find from the relation between current, voltage and power, that the current in the transmission line will be equal to 5 amperes instead of 100 amperes. It will then be possible to have a 200 volt drop in the line and still have a voltage of 2000 on the primary of the transformer at the distant end, or the required 100 volts when stepped down. Since the current in the line will now be 1/20th of 100, or 5 amperes, the resistance of the line in this case will be—

$$R = \frac{200}{5} = 40 \text{ ohms}$$

We find, then, that the size of the conductors for the transmission line where the transformers are used, must be such that the resistance will not exceed 40 ohms, whereas in the first case it must be such that the resistance will not exceed 1/10th ohm. The amount of copper required in the second case is 1/400th or only 1/4th of one per cent of that required in the first case. The economy due to the copper saving is apparent.

14.3 Transformer Applications to Telephone Circuits

The applications of transformers to telephone circuits are numerous and varied. The reduction of energy losses in alternating-current transmission, as illustrated in Figure 14-6, has an application to telephone transmission but is not so important as other uses. One very general use is to accomplish the result given as b above. In this case, the primary function of the transformer is to transfer energy to another circuit rather than to change the voltage and current values. When

so used in telephone work, they are generally called **repeating coils** rather than transformers because their primary function is to "repeat" the energy into a different circuit rather than to transform it into a different state. There are, however, inequality ratio repeating coils which perform both functions. On the other hand, in connection with telephone repeater circuits and certain other telephone apparatus, input and output coils are used primarily to match impedances to permit maximum energy transfer, as explained in later Chapters.

One of the most common applications of the repeating coil in telephone work is in connection with the common battery cord circuit, as illustrated by Figure 14-7. Here the alternating-current flow in one subscriber's line is repeated
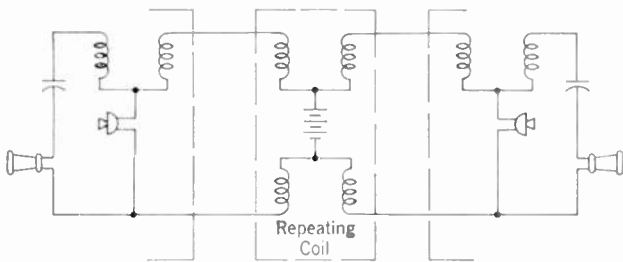

FIG. 14-7 TRANSFORMERS IN COMMON BATTERY TELEPHONE CONNECTION

into the other subscriber's line with little energy loss, and at the same time the windings of the coils afford the proper direct-current connections for each subscriber's station to receive a super-posed d-c current for transmitter supply. Another very general use of repeating coils in the telephone plant is for deriving "phantom" circuits. Here the coils serve a unique purpose which has no counterpart in electric power work, and is not included in the classification of transformer functions given above. We shall therefore need to consider this application more fully. However, it may be noted that the coils, while serving this particular purpose, may also function as impedance matching devices.

14.4 The Phantom Circuit

Figure 14-8 is a simplified diagram of two adjacent and similar telephone circuits arranged for phantom operation. By means of repeating coils installed at the terminals of the wire circuits, a third telephone circuit is obtained. This third cir-

cuit is known as the phantom and utilizes the two conductors of each of the two principal, or "side" circuits, as one conductor of the third circuit. The two side circuits and the phantom circuit are together known as a phantom group. These three circuits, employing only four line conductors, can be used simultaneously without interference with each other, or without crosstalk between any combination, provided the four wires have identical electrical characteristics and are properly "transposed" to prevent crosstalk.

The repeating coils employed at the terminals are designed for voice-current and ringing-current frequencies, and do not appreciably impair transmission over the principal or side circuits. The third or phantom circuit is formed by connecting to the middle points of the line sides of the repeating coil windings, as shown in the Figure. Since the two wires of each side circuit are identical, any current set up in the phantom circuit will divide equally at the mid-point of the repeating coil line windings. One part of the current will flow through one-half of the line winding, and the other part of the current will flow in the opposite direction through the other half of the line winding. The inductive effects will be neutralized, and there will be no resultant current set up in the drop or switchboard side of the repeating coil. Since the phantom current divides into two equal parts, the halves will flow in the same direction through the respective conductors of one side circuit, and likewise return in the other side circuit. At any one point along a side circuit, there will be no difference of potential between the two wires due to current in the phantom circuit, and a telephone receiver bridged across them will not detect the phantom conversation.

Since there is no connection, inductive or otherwise, between the two circuits at the terminals, it is equally true that a conversation over a side circuit cannot be heard in the phantom. This can
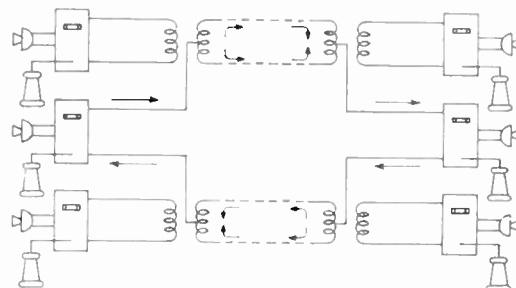

FIG. 14-8 PRINCIPLE OF THE PHANTOM CIRCUIT

be understood by imagining a flow in the closed side circuit through the line wires and the windings of the repeating coils at each end. With the side circuit conductors electrically equal, there can be no difference of potential between the mid-point of the repeating coil line winding at one end and the mid-point of the repeating coil line winding at the other end because the drops of potential for the two parts of the side circuits are equal and opposite. If the side circuit, therefore, impresses no difference of potential on any part of the phantom circuit, the side circuit conversation cannot be heard over the phantom.

In the theory of the phantom it should not be forgotten that the conductors are assumed to be electrically identical, or in other words, the conductors are perfectly "balanced". The phantom is very sensitive to the slightest upset of this balance, and circuits that are sufficiently balanced to prevent objectionable crosstalk or noise in physical circuit operation, may not be sufficiently balanced for successful phantom operation.

14.5 Standard Repeating Coils

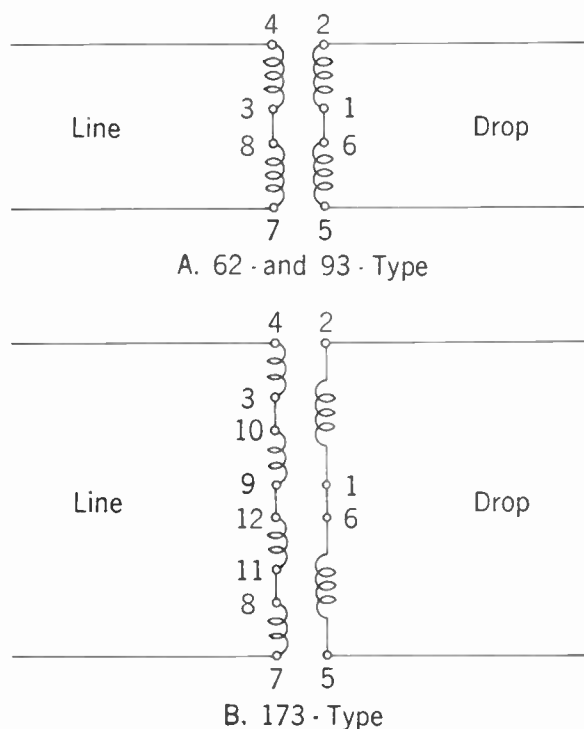A number of general types of repeating coils are currently standard in the Bell System. One

A. 62 - and 93 - Type

B. 173 - Type

Fig. 14-9 Standard Repeating Coils

TABLE VII
STANDARD REPEATING COILS

IMPEDANCE RATIO LINE TO DROP 4–3 AND 8–7 TO 2–1 AND 6–5	93-TYPE	62-TYPE
1:1	93-A	62-A
1:1.62	93-B	62-B
1.62:1	93-F	62-C
2.66:1	93-G	62-E
1.24:1	93-H	
2.28:1	93-J	
1:1.28		62-F
1:2.34		62-G

principal type, illustrated by the 62 and 93 series, has four windings, the terminals of which are designated by numbers as shown by Figure 14-9 (A). The other type, illustrated by the 173 series, has six windings which may be connected as shown in Figure 14-9 (B) with four windings on the line side, or with the 9-10, 11-12 windings not used, depending on the impedance ratio required. In all types, the windings which are used to form the line side are precision manufactured so as to be as nearly identical electrically as possible. This balance is required, as we have already seen, to avoid crosstalk where the coils are used in phantom operation. The drop windings (that is, 1-2 and 5-6) do not need to be so well balanced in normal use.

The 62- and 93-type coils have toroidal cores made of many turns of fine-gage silicon-steel wire sawed through at one point to introduce a gap in the magnetic circuit. In the 93-type coil this gap is filled with compressed powdered iron which, while increasing slightly the reluctance of the core gives it a high degree of magnetic stability, preventing permanent magnetization under abnormal service conditions. In the 62-type coil the gap in the magnetic circuit is unfilled which tends to make the coil even more stable. This coil is especially well adapted for use on circuits composited for d-c telegraph operation. The same feature, however, tends to make the 62 series inefficient at low frequencies and they cannot be used on circuits employing 20-cycle signaling, whereas the 93 series may be used for such purposes. Standard 173-type coils are built with permalloy cores.

The types of repeating coils discussed above are manufactured with a number of different turn ratios to provide various impedance matching

combinations. Table VII gives the standard impedance ratios for 93- and 62-type coils. The 173-type coils are likewise available in a wide range of impedance ratios. The impedance ratio obtained in their use depends on whether all four of the line windings are used and on how those used are connected. The impedance ratios that can be obtained accordingly do not lend themselves readily to tabular presentation, but various ratios, line-to-drop, ranging from as low as 0.6:1 to as high as 2.52:1 may be obtained.

14.6 Radio-frequency Transformers

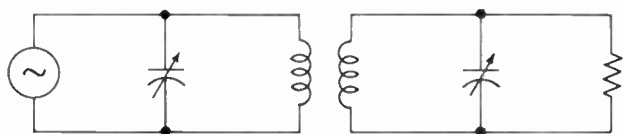At the relatively high frequencies employed in radio circuits, it is impractical to use transform-

ers having iron or other ferromagnetic cores because the losses resulting from circulating or eddy currents in the metallic cores become excessive. For frequencies up to about 1,000,000 cycles (1 mc), however, it is possible to use cores made of the magnetic materials known as **ferrites**. These are ceramic-type materials formed by compressing powdered mixtures of the oxides of various metals, such as nickel, zinc, magnesium, manganese, and iron, into the desired shape and "firing" at temperatures in the order of 1000° C. Electrically, the ferrite core is a semi-conductor and its resistivity may be approximately one million times as high as that of the ordinary iron core. This reduces eddy current losses to negligible values up to very high frequencies. At the same time, the ferrite's magnetic permeability is high up to several megacycles.

Before the development of ferrite materials it was generally necessary to use air core transformers for most radio applications and these are still widely employed in ordinary radio circuits. It is not possible, however, to build such a transformer so that all of the flux lines set up by one coil will cut all of the turns of the other coil because the reluctance of air to the flow of magnetic flux is so high. The simple voltage, current and impedance relationships that apply to iron-core transformers therefore no longer hold true. In practice, air-core transformers are designed largely on an empirical or "cut and try" basis and large step-ups of voltage cannot ordinarily be obtained.

When air-core transformers are used for coupling between radio frequency or other high-frequency circuits, the primary, secondary, or both coils, are often tuned to resonate at the operating frequency. Such a circuit is shown in Figure 14-10 where both primary and secondary are tuned by means of variable capacitors. The tuning is effected by adjusting the values of capacitance so as to satisfy the resonance equation—

$$ f = \frac{1}{2\pi\sqrt{LC}} $$

In this way parallel-resonant circuits of high impedance are obtained. In addition, a circuit of this sort will act as a filter and pass a band of frequencies the width of which is determined by the Q of the circuit.

14.7 The Hybrid Transformer

There are a number of situations in telephone practice where it is desirable or necessary to employ separate circuits for transmission in opposite directions. Thus, an ordinary two-wire voice-frequency telephone circuit readily transmits electrical energy representing the voice message in either direction. When a telephone repeater of the 22-type (Chapter 25) is inserted in such a circuit to counteract the energy losses along the line, however, it is necessary to separate the circuit into two one-way paths. This is because the amplifiers used in the repeater are inherently unidirectional devices.

A method for effecting such a separation is illustrated in Figure 14-11. Here, each end of the opened two-wire line is connected to a special type of transformer, known as a **hybrid coil** or **three winding transformer**. Each hybrid coil has six
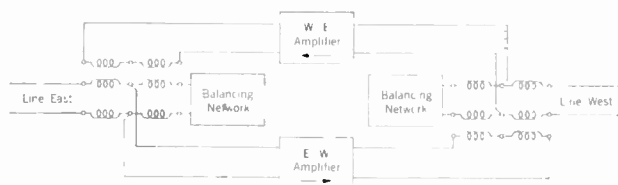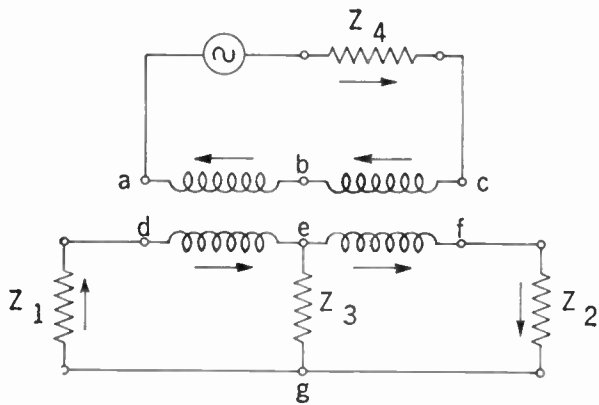
FIGURE 14-12

windings, as indicated, and all six windings are wound on a single magnetic core. The circuit arrangement also includes two balancing networks, each designed ideally to have exactly the same impedance as the line section which is connected to the same coil. The function of the overall circuit arrangement is to permit energy entering the coil from the East line to pass to the E-W amplifier and thence out on the West line through the West coil. Similarly energy coming from the West line passes through the W-E amplifier and out on the line East. At the same time, the circuit must prevent the output energy of either amplifier from crossing a coil to enter the input of the other amplifier, since this would set up a local circulation that would cause the repeater to "sing" or "howl".

How the hybrid coil acts to meet those requirements may be understood by an analysis of the simplified circuit of Figure 14-12. This represents a single coil connected to four impedances. Z_1 is the impedance of the line and Z_2 is the equal impedance of the balancing network. Z_4 is the impedance of one amplifier output and Z_3 is that of the other amplifier input. Two of the line windings of the coil are omitted for simplicity since their presence would have no bearing on the analysis.

Consider first a source of voltage inserted in series with the amplifier output Z_4. This establishes a current through Z_4 and the identical coil windings ab and bc, as indicated by the arrows. This induces voltages across de and ef which are exactly equal in value because these two windings are also identical. The resultant current will produce equal voltage drops across Z_1 and Z_2, because these two impedances have equal values. The potential at e is accordingly the same as at g and no current will flow in Z_3. In other words, there is

no transmission from Z_4, the output of one amplifier to Z_3, the input of the other amplifier. It may be noted, however, that only half of the energy delivered from Z_4 goes into the line Z_1, the other half being dissipated in the balancing network Z_2. The amplifier must therefore be adjusted to supply twice as much energy as it is desired to feed to the line.

Transmission through the coil in the other direction may be studied with the help of Figure 14-13 where the voltage source is shown in series with the line impedance Z_1. The behavior in this case is best followed by first assuming that Z_2 is disconnected, leaving the terminals fg open. The voltage source then sets up a current through Z_1, Z_3 and the winding de, as shown by the arrows. The current through de induces a current in winding ab which also flows through Z_4 and winding bc. This in turn induces a voltage across ef which is the same in value as that across de since these two windings are identical. The turn ratio of the transformer is such that the impedance Z_4, as seen through the winding de, is equal to Z_3. The potential difference between d and e is then equal to the drop between e and g across the impedance Z_3. It is also equal to the potential difference between e and f, as noted above. But the voltage across ef is opposite in direction to the drop across Z_3 so that points f and g are at the same potential. Therefore, no current will flow in Z_2 if it is now reconnected. This means that energy coming from the line divides equally between Z_3 and Z_4 and none reaches the balancing network Z_2. Again, however, only half of the incoming energy reaches the input of the amplifier where it is useful, the other half being dissipated in the output of the other amplifier. The amplifier must be designed to counteract this loss also.
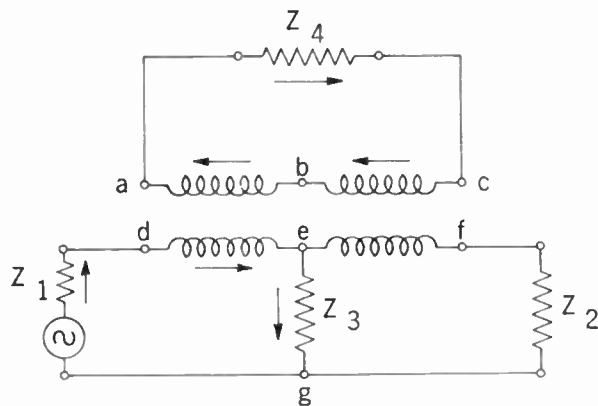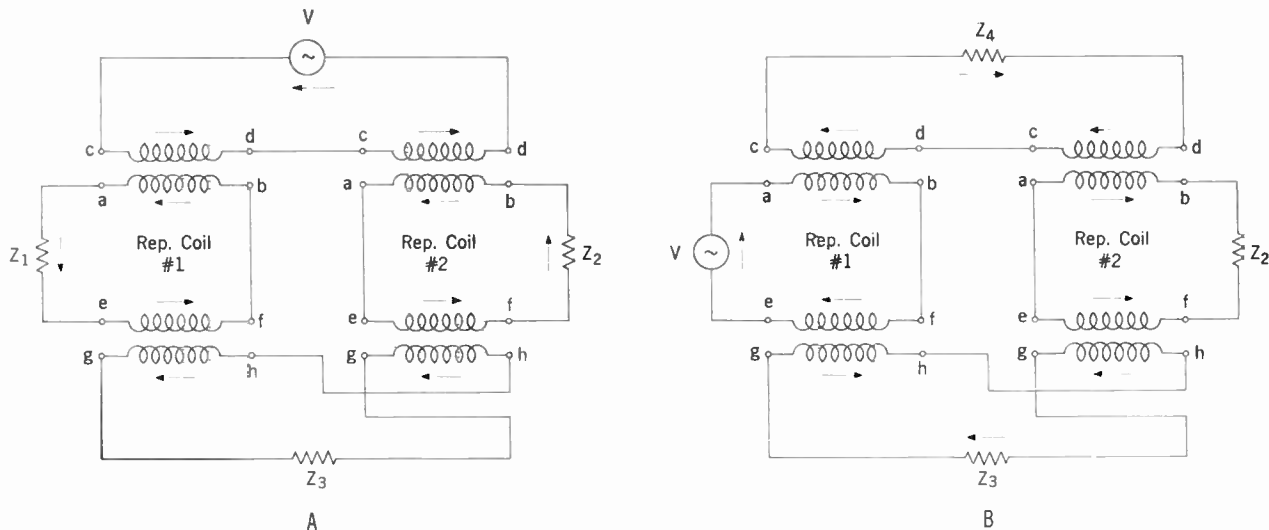

FIGURE 14-13

FIG. 14-14 FOUR-WIRE TERMINATING SET

Hybrid coils of the type discussed above can also be used at the terminals of 4-wire telephone circuits to convert the 4-wire line into a 2-wire line. In practice, however, a slightly different transformer arrangement, known as a 4-wire terminating set is more generally used for this purpose. This consists of two ordinary repeating coils connected with one winding of one coil reversed, as shown in Figure 14-14. Such an arrangement can also be used with the 22-type telephone repeater if desired.

The principle involved here is the same as for the hybrid coil proper, as may be seen by analyzing the circuit. Thus, we may consider first the case of energy coming from the transmitting side of the 4-wire line for transfer to the 2-wire line. This is illustrated by Figure 14-14(A) where the energy source is represented by V, the 2-wire line impedance by Z_1, the equal balancing network impedance by Z_2, and the impedance of the receiving side of the 4-wire line by Z_3. As indicated by the arrows, at any given instant V sets up equal voltages in Z_1 and Z_2, but because the winding g-h of repeating coil 2 is reversed, the voltage set up in this winding is opposed by the equal voltage set up in winding g-h of repeating coil 1. As a result, no current is established in Z_3. Similarly, where the energy comes from the 2-wire line, as illustrated in Figure 14-14 (B), equal voltages are set up in Z_3 and Z_4 and there is no current in the network, Z_2. This is because the direction of the voltage set up in winding e-f of repeating coil 2 is such as to oppose the equal voltage set up in winding a-b.

Where repeating coils having six windings are used in a hybrid arrangement, the connections are naturally somewhat different. Thus Figure 14-15 shows the arrangement at one side of a repeater for a two-wire circuit. Here it is assumed that the line and balancing network impedances are equal and that the impedances connected at HYB OUT and HYB IN are equal.

Current received from the line flows through the line windings of both coils and induces equal voltages in the network windings (9-10 and 11-12) of both coils. Because of the reversed poling of the network windings of coil (B) with respect to those of coil (A) the resultant voltage across the network is zero. Equal voltages are also induced in the 2-1 and 6-5 windings of both coils and the power received from the line divides equally between the impedances connected at HYB OUT and HYB IN.

Current received from the output of one ampli-
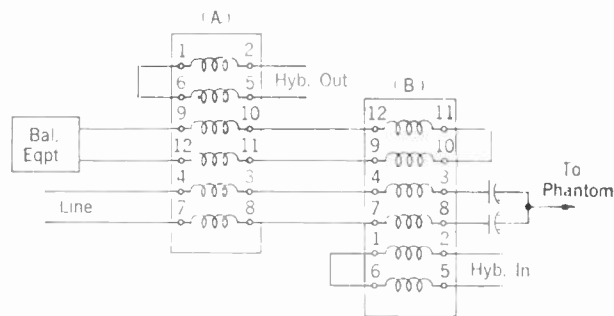
FIG. 14-15 REPEATING COIL HYBRID ARRANGEMENTS FOR 2-WIRE CIRCUITS

[129]

fier at HYB IN flows through windings 2-1 and 6-5 of the (B) coil and induces equal voltages in the line and network windings. The currents in these windings flow through the corresponding windings of the (A) coil to line and balancing network respectively. The currents in the line and network windings of coil (A) induce equal voltages in the 2-1 and 6-5 windings but these voltages are opposite in phase because of the re-versed poling of the network windings, and the resultant voltage is zero. The power received from the amplifier divides equally between the line and the balancing network.

This arrangement requires that the windings of a given coil be mutually balanced to a high degree of precision in order to secure good trans-hybrid balance but does not require a high degree of balance between separate coils.

CHAPTER 15

TELEPHONE TRANSMISSION THEORY

15.1 The Transmission System

Any transmission system consists of three essential parts; a source of energy, a medium over which it is desired to transmit energy to a receiving device, and the receiving device itself, which usually converts the electric energy into some form more useful. In a power transmission line, an electric generator may be the source of energy; high voltage lines with transformers at either end may be the transmitting medium; a motor, lamp, or heater may be the receiving device for converting electric energy into some other useful form. In a long distance telephone connection, a transmitter may be considered as the source of energy; the line from the speaking party to the listening party with all of its associated conductors, coils, and connections, may be thought of as the transmission medium; and the telephone receiver at the distant end may be considered as the third part of the transmission system, or the device which converts small electric currents into audible vibrations of air called sound waves.

In this and the following two Chapters we shall be concerned with the transmission line or other medium over which energy is carried from the transmitting to the receiving device. Our study will begin with an analysis of the characteristics of ordinary wire conductors at what we may call relatively low frequencies—i.e., voice frequencies and carrier frequencies ranging up to approximately 150 kc. Later we shall consider transmission at higher frequencies, including radio frequencies extending up to several thousand megacycles. This will require us to make some study of certain types of transmission mediums other than the usual wire line. Before beginning our analysis of the properties of the transmission line, however, we may devote some attention to certain basic theorems regarding networks and power transfer that will be useful in the subsequent analysis.

15.2 Transfer of Power

If a transmission system is to accomplish its purpose, it must be so designed that the receiving device is successfully operated as a result of the application of energy at the source. As a secondary consideration it may be designed for power efficiency—that is, regardless of the magnitude of the power delivered to the receiving device, to keep at a minimum the power lost in transmitting the energy from the source. Although this is important in any transmission system, its special importance is in power transmission. In telephone work we probably think more of the primary purpose, that is, the system's effectiveness in operating the receiving device, regardless of what percentage of power may be lost.

The power efficiency of a transmission system may be determined by dividing the received power by the power delivered to the system at the generator end. This efficiency will increase as the resistance of the receiving device increases, and the most efficient operation is accordingly obtained when the receiving circuit has high resistance. In most telephone applications, on the other hand, we are interested in receiving all the power possible regardless of the efficiency. In a system containing only resistance the condition for maximum transfer of power is obtained when the resistance of the receiving circuit is equal to the resistance of the network to which it is connected, as measured across the receiving terminals. The simplest application of this is secured by connecting to a battery a resistance equal in magnitude to the internal resistance of the battery. In this case, the battery will transfer to the external circuit the maximum amount of power, but in doing so
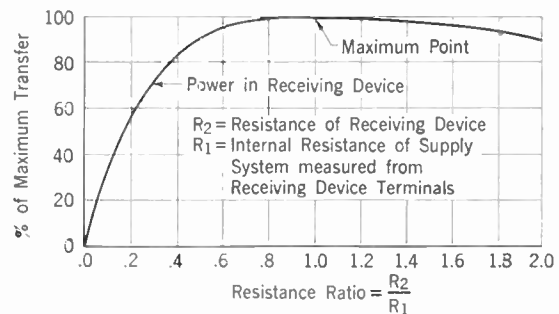

FIG. 15-1 POWER TRANSFER AS A FUNCTION OF RESISTANCE RATIO
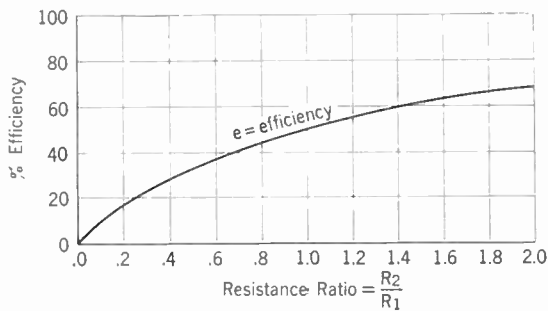
[131]

FIG. 15-2 EFFICIENCY AS A FUNCTION OF RESISTANCE RATIO

will operate at an efficiency of only 50 per cent.

Figure 15-1 shows a curve which represents the power in a receiving circuit for various values of its resistance expressed as a ratio of the fixed resistance of the supplying circuit. Figure 15-2 shows the efficiency for the same conditions.

Maximum power transfer in a-c circuits, where reactances are involved, is obtained in the general case when the impedance of the receiving device is conjugate to the impedance looking back into the network—that is, when the resistances are equal and the reactances are equal and opposite in sign.

15.3 Thevenin's Theorem

For the purpose of simplifying electrical calculations, we can consider any electric system as one network supplying energy to another. The first of these networks may then be replaced by an equivalent simplified circuit, which consists of an emf and an impedance in series.

This means that regardless of how complicated an electric circuit may be, its effect in supplying current to any other circuit connected to it at two designated terminals, is equivalent to some source of electromotive force in series with an impedance. This principle is called Thevenin's Theorem and Figure 15-3 illustrates its use. Here E is a source of electromotive force connected to a complicated network. If it is desired to connect some receiv-
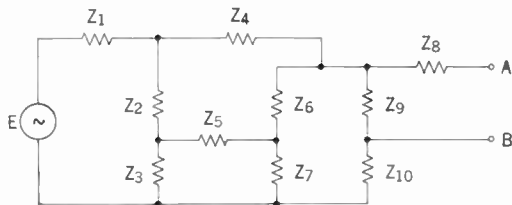

FIGURE 15-3

ing device to terminals A and B, the effect of this electric system on the receiving device will be the same as that of the electric system shown by Figure 15-4 where E' is the voltage between terminals A and B of Figure 15-3, and R' is the impedance measured or calculated from the same terminals with the electromotive force E considered as having zero value. Thevenin's Theorem may be briefly stated as follows:

The current supplied to a receiving device connected to two terminals of any electrical system is equal to the voltage between these terminals before the device is connected, divided by the im-
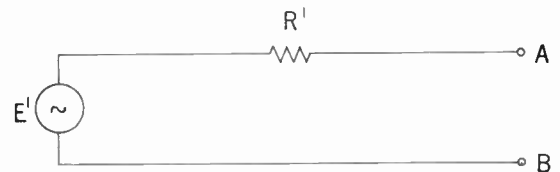
FIG. 15-4 APPLICATION OF THEVENIN'S THEOREM TO THE NETWORK OF FIG. 15-3

pedance of the receiving device plus the impedance measured or calculated across the terminals with any sources of emf considered as having zero value.

15.4 Equivalent Networks

Thevenin's Theorem gives us a method of substituting a simple equivalent circuit for any complicated electric system, but in so doing we are required to replace the source of electromotive force with one having another value. It is often

FIGURE 15-5

desired to determine the simplest equivalent network for a complicated electric system, which will supply to some receiving device the same current as the electric system and will take from the same source of electromotive force the same current as the electric system. A network consisting of three impedances of proper value arranged in the form

[132]

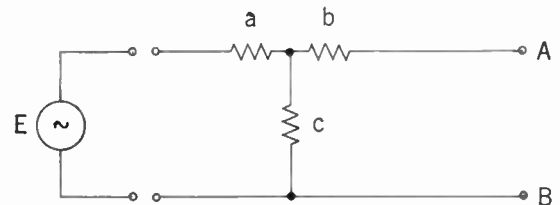of a **T** as shown by Figure 15-5 may be substituted for such a complex network as that shown by Figure 15-3, and the current supplied to this system by the electromotive force E will remain unchanged, and the current received by a device connected to the terminals A and B will be the same. As we shall see in a later Chapter, this same result can also be effected by means of a simple network having three arms arranged in the form of a π.

In determining values for the three impedances in an equivalent T-network such as is shown by Figure 15-5, the following equations may be used:

$$\text{Impedance of } a = Z_1 - c \qquad (15{:}1)$$

$$b = Z_3 - c \qquad (15{:}2)$$

$$c = \sqrt{(Z_1 - Z_2)Z_3} \qquad (15{:}3)$$

where Z_1 is the impedance of the complicated network at the terminals connected to the source of emf with the receiving device terminals open; Z_2 is the same with the receiving device terminals short-circuited; and Z_3 is the impedance of the complicated network at the receiving device terminals with the source terminals open.

At a single frequency, a long transmission line can be exactly represented electrically by a simple three-element equivalent network, although the determination of the values of the three arms involves in this case the use of certain higher branches of mathematics. For many purposes, it is convenient to deal with a long transmission line by considering it as being made up of a number of separate T-sections, each identical and each representing approximately a short section of the transmission line. Obviously the degree of approximation will be higher the larger the number of sections used to represent a given line. Figure 15-6 illustrates such a multisection uniform network where each T-section consists of two series arms of value $Z_1/2$ and a shunt arm Z_2.

15.5 Characteristic Impedance

One thing that we will ordinarily wish to know about a transmission line, or a multisection uniform network representing such a line, is the input impedance of the line. If a line is very long, this input impedance will be found to depend entirely upon the characteristics of the line itself and will not be affected by what, if anything, is connected to the far end of the line.

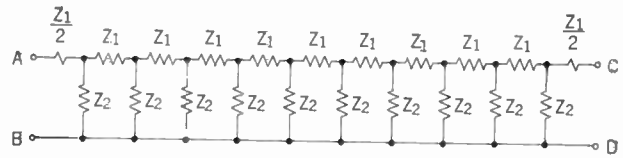If we should measure or calculate the impedance of the network of Figure 15-6 across the input terminals A and B, we would obtain a certain impedance value which we may call Z'_0. Now let us connect this impedance Z'_0 across the receiving terminals C, D. As seen from the sending end, this will make the network 20 sections long instead of 10 sections long. If we then repeat this process, we will arrive at a new impedance which connected across the terminals C, D will make the network 30 sections long. Very soon we will arrive at an impedance value which will not change: in other words, the network will become infinitely long for all practical purposes. This impedance is the **characteristic impedance** Z_0 of the network. That Z_0 depends only on the impedance values of the unit sections of the network is obvious from the fact that those were the only impedance values used in determining it. This being so, it follows that if we take one section of the network and
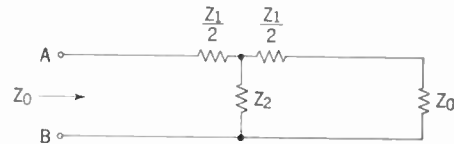
FIGURE 15-7

terminate it with its characteristic impedance Z_0 as shown in Figure 15-7, the input impedance as shown across A and B will also be Z_0. Z_0 is then readily calculated from the Figure as follows:

$$Z_0 = \frac{Z_1}{2} + \frac{Z_2\left(\dfrac{Z_1}{2} + Z_0\right)}{\dfrac{Z_1}{2} + Z_2 + Z_0}$$

$$\frac{Z_0 Z_1}{2} + Z_0 Z_2 + Z_0^2 = \frac{Z_1^2}{4} + \frac{Z_1 Z_2}{2} + \frac{Z_0 Z_1}{2} + \frac{Z_1 Z_2}{2} + Z_0 Z_2$$

$$Z_0^2 = \frac{Z_1^2}{4} + Z_1 Z_2$$

$$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1 Z_2} \qquad (15{:}4)$$

As has already been pointed out, the more identical T-sections that are used in the multisection network to represent a given length of transmis-

FIG. 15-6 MULTISECTION UNIFORM NETWORK

FIGURE 15-8

sion line, the more closely does the network approximate the line. If we take a transmission line having a series impedance Z_1 per mile and a shunt impedance Z_2 per mile and represent each mile by n T-network sections, the impedance of each series T-arm will be $Z_1/2n$ and each shunt arm will be nZ_2. The characteristic impedance of the network may then be written from equation (15:4) as—
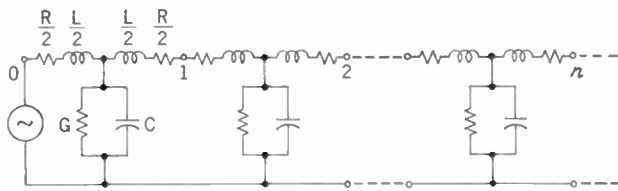
$$Z_0 = \sqrt{\frac{Z_1^2}{4n^2} + \frac{Z_1}{n} \, nZ_2}$$

From this it may be seen that as the number of sections n is increased to a very large number, the first term under the radical approaches zero and the equation reduces to—

$$Z_0 = \sqrt{Z_1 Z_2} \qquad (15:5)$$

Actually the use of higher mathematics, as previously mentioned, will show that this is the exact expression for the characteristic impedance of a uniform transmission line.

The fundamental transmission characteristics of the wire transmission line are determined by four basic properties known as the "primary constants" of the line. These are the series resistance of the conductors, R; the inductance of the conductors, L; the capacitance between the two conductors, C; and the leakage resistance or conductance between the two conductors, G. R and L are series constants and C and G are shunt constants. In terms of these constants, a multisection uniform network simulating the transmission line may be represented as shown by Figure 15-8. The series impedance of a single section of this network may be written as—

$$Z_1 = R - j\omega L \qquad (15:6)$$

where the symbol ω is used for convenience to represent $2\pi f$. The impedance of the parallel shunt branch similarly is—

$$Z_2 = \frac{1}{G + j\omega C} \qquad (15:7)$$

From equation (15:5) the characteristic imped-

ance may then be written as—

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \qquad (15:8)$$

Here R is resistance in ohms per unit length of line; L is inductance in henries; G is conductance in mhos; C is capacitance in farads.

As the above equation shows, the value of characteristic impedance is dependent solely on the primary constants of the line and the frequency. At any given frequency, therefore, this impedance has a fixed value for any given type of line regardless of the length of the line or what may be connected to the line terminal. One of its most useful applications in practical telephone work lies in the fact that receiving devices to be connected to a line may be designed with impedances equal to the characteristic impedance of the line, thus permitting a maximum transfer of power out of the line.

15.6 Propagation Constant

If a transmission line having characteristic impedance Z_0 is infinitely long, or is terminated at the receiving end with an impedance equal to its characteristic impedance, the current entering the line when a voltage E is applied at the sending end will obviously be $I_0 = E/Z_0$. Our next main interest in the line will be to know what part of this current or applied energy will reach the receiving device. We may also be interested in knowing how long it will take it to get there. To answer these questions we may again simulate the line by a multisection network and analyze what happens in each individual section. Clearly, since there is a shunt path across the line in each section, the current leaving a section will be less than that entering it in the direction of transmission. But since the sections are all identical, the ratio of leaving to entering current will be the same for all sections; that is—

$$\frac{I_1}{I_0} = \frac{I_2}{I_1} = \frac{I_3}{I_2} \text{ etc.}$$

Accordingly, we may write generally for the current at the end of the nth section—

$$\frac{I_n}{I_0} = \frac{I_1}{I_0} \times \frac{I_2}{I_1} \times \frac{I_3}{I_2} \times \cdots \frac{I_n}{I_{n-1}}$$

or

$$\frac{I_n}{I_0} = \left[\frac{I_n}{I_{n-1}} \right]^n \qquad (15:9)$$

[134]

An equation such as (15:9) is most conveniently dealt with by means of logarithms. Expressed in this way it may be written—

$$\log_e \frac{I_n}{I_0} = -n\gamma \text{ or } 2.303 \log \frac{I_n}{I_0} = -n\gamma \quad (15:10)$$

where γ is **defined** as the **propagation constant** of the transmission line. Its value is dependent on the primary constants of the line and the frequency. It may be determined to any desired degree of approximation by an analysis of the current relationships in a single network section terminated in Z_0 as illustrated by Figure 15-7. Its exact value will thus be found to be—

$$\gamma = \sqrt{\frac{Z_1}{Z_2}} = \sqrt{(R + j\omega L) \div \frac{1}{G + j\omega C}}$$
$$= \sqrt{(R - j\omega L)(G + j\omega C)} \quad (15:11)$$

The propagation constant γ is a vector quantity because it is equal to $\sqrt{Z_1/Z_2}$ and both Z_1 and Z_2 are vector quantities. It may accordingly be separated into a "real" and "imaginary" component thus—

$$\gamma = \alpha + j\beta \quad (15:12)$$

where α, the real part of γ, is a measure of the amount by which the current or voltage is reduced in each unit length of line and is called the **attenuation constant.** Similarly β, the imaginary part of γ, provides a measure of the phase shift, and is related to the time required for the energy to traverse the unit length of line. It is called the **phase or wavelength constant.** It should be noted that while β represents an angle of phase shift, it should not be confused with any phase angle which may exist between the current and voltage at a particular point on a line. In practice it is usually easier to evaluate α and β by making use of equations (15:11) and (15:12). They may be expressed, however, directly in terms of the primary constants—R, L, G and C, as follows:

$$\alpha = \sqrt{\tfrac{1}{2}\sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} + \tfrac{1}{2}(GR - \omega^2 LC)} \quad (15:13)$$

$$\beta = \sqrt{\tfrac{1}{2}\sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} - \tfrac{1}{2}(GR - \omega^2 LC)} \quad (15:14)$$

The value of β gives directly in radians the phase shift in current or voltage over a unit length of line, such as one mile, having primary constants R, L, G and C. (A radian is $360/2\pi$ or 57.3 degrees.) For a length of line comprising n unit lengths the total phase shift will obviously be $n\beta$ radians. The value of α may be used to determine the current or voltage attenuation by applying the real part of equation (15:10). Thus, for a line n units in length—

$$-n\alpha = 2.303 \log \frac{I_n}{I_0} = 2.303 \log \frac{E_n}{E_0} \quad (15:15)$$

where the current and voltage ratios express magnitudes only without regard to any phase differences that may exist. The power entering the line is—

$$P_a = E_0 I_0 \cos \theta$$

and the power at the end of the nth unit length is—

$$P_n = E_n I_n \cos \theta$$

θ in both cases being the angle of the characteristic impedance of the line. The magnitude of the power loss or attenuation may then be determined from—

$$2.303 \log \frac{P_n}{P_0} = -2n\alpha \quad (15:16)$$

The power, therefore, is seen to die out or attenuate in a ratio which is the square of the current and voltage ratio.

In the foregoing we find for the most part a mathematical significance of α and β. Let us now analyze the physical circuit to determine what actually happens as the current is sent from point to point. In order to simplify the analysis, we shall start with an actual cycle of emf impressed on the sending end of a multisection network, and
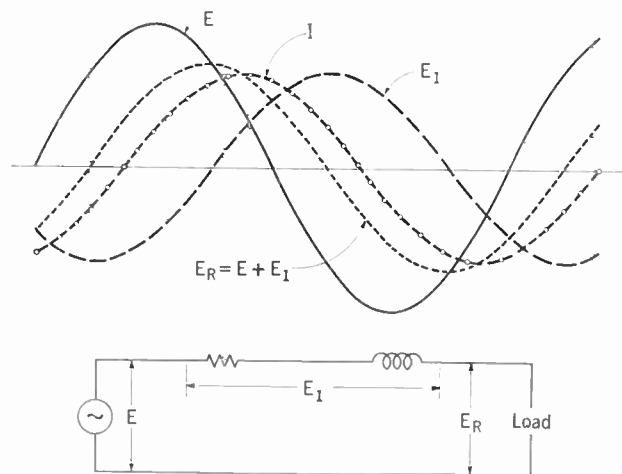

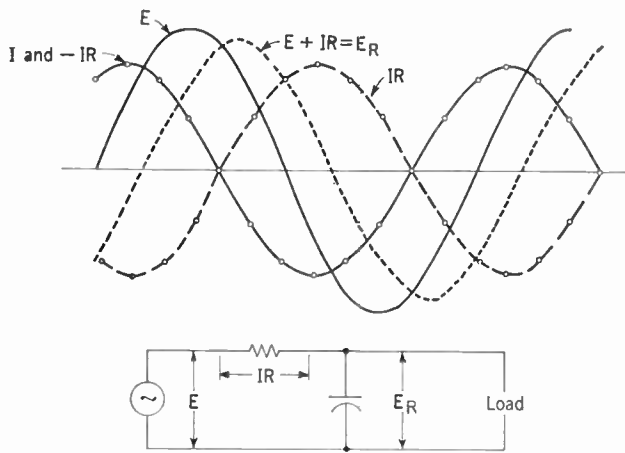FIG. 15-9 VOLTAGES AND CURRENT IN AN INDUCTIVE CIRCUIT

[135]

FIG. 15-10 VOLTAGES AND CURRENT IN A CAPACITIVE CIRCUIT

consider separately the effects of inductance and capacitance on the propagation.

From our previous study, we know that inductance acts to cause the current to lag behind the impressed voltage, so that in a circuit made up of resistance and inductance we would expect a lagging current. Figure 15-9 shows the time relationship between voltage and current in such a circuit, where E is the voltage curve, and I the current curve. This current sets up a back or induced emf E_1, which is the sum of the IR drop across the resistance and the IX drop across the

inductance. It combines with the original voltage E to give the resultant voltage E_R on the load side of the inductance. The curve E_R is obtained by adding E and E_1 and it will be observed that the resulting curve lags behind E, the original voltage. A circuit containing resistance and capacitance, on the other hand, produces a leading current as shown by Figure 15-10, and this current produces an IR drop which is opposite in phase to the current. Now if we combine the IR drop and the voltage, we obtain the resultant voltage E_R, which exists across the capacitor and the load. This voltage likewise lags behind E, the original voltage.

In both cases we have obtained a resultant voltage which lags behind the impressed voltage. Bridged capacitance assists series inductance in the phase retarding effect. Due to the presence of reactance, therefore, the voltage has been "delayed in time", so that the maximum voltage reaches any given point along the network later than it would if the reactance were not present. Each section of the network, due to resistance and leakage, absorbs energy and thereby reduces the voltage which can act on the next section. Further, the voltage available at the next section lags behind the voltage impressed on the section, so that as we move away from the generator, the acting voltages are lagging farther and farther behind
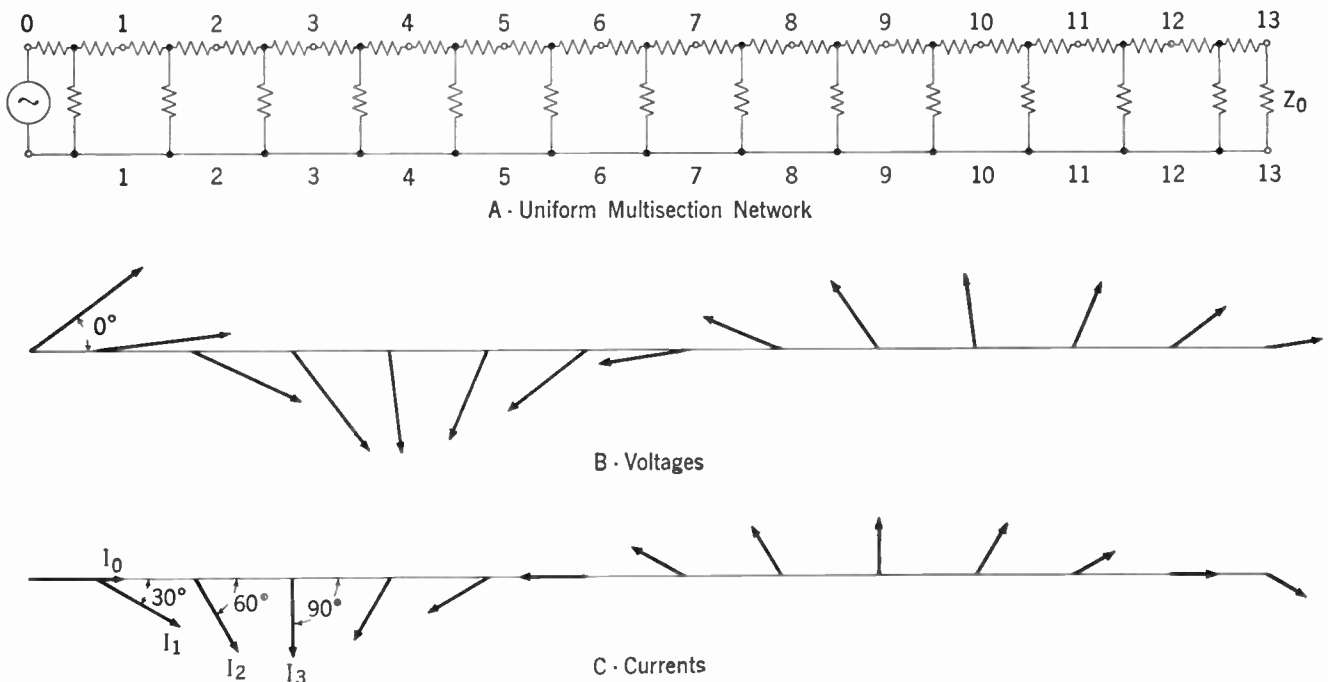

A · Uniform Multisection Network

B · Voltages

C · Currents

FIG. 15-11 TRANSMITTED CURRENTS AND VOLTAGES AT JUNCTIONS OF A MULTISECTION UNIFORM NETWORK

[136]

the generator voltage. Here we have a connecting link between **time** and geographical **distance** traveled along the line.

To bring this out clearly, let us assume that we take our sections of such a length that, for a frequency of 1000 cycles, the time lag between voltages can be represented by 30 degrees per section on the **time-voltage** diagram; if we simulate by each section fourteen and three-quarters miles of 104 open wire circuit, we will obtain such a relationship. In order to make the story complete, we will also assume the reduction in voltage magnitude due to resistance and leakage loss to be such as to give a ratio of 0.895 between the end and beginning of each section. If we assume the original voltage E_0 to be 10 volts, the voltage at the end of the first section, E_1, will be 8.95 volts, lagging 30° behind E_0. E_2, at the end of the second section, will be 0.895×8.95 or 8.01 volts, lagging 30° behind E_1 or 60° behind E_0. If we represent the voltages at various points by vectors, we will obtain a system of vectors as shown in Figure 15-11(B), where the multisection network is shown as Figure 15-11(A) and the voltage acting at each junction is directly below.

Since the ratio of current to voltage is constant, it follows that the chart representing currents will have the same form, with each vector proportional and removed by an angle θ from the corresponding voltage vector, where θ is the angle of the characteristic impedance Z_0. Thus we may treat a similar Figure such as 15-11(C) as a "distance-current diagram" where the vectors, I_0, I_1, I_2, etc., show the magnitude and relative phase of the currents at the network junctions. If now we refer all the current vectors to a common reference point, we will obtain a broken curve such as that of Figure 15-12(A), which shows graphically how the currents at various points are related. In this Figure the vector $I_0 = G\text{-}0$ is the current entering the first section and $I_1 = G\text{-}1$, the current leaving that section. Then the vector 1-0 must be the current that passes through the shunt in the first section, because the sum of the current through the shunt and the current going ahead gives 1-0 as the resultant of the vector diagram. This is perhaps more clearly illustrated by Figure 15-12(B). For the same reason 2-1 will be the current passing through the second shunt, etc.

We may, therefore, conceive of the total entering current as the resultant of a number of component currents which flow from the generator through the various shunt paths and back to the
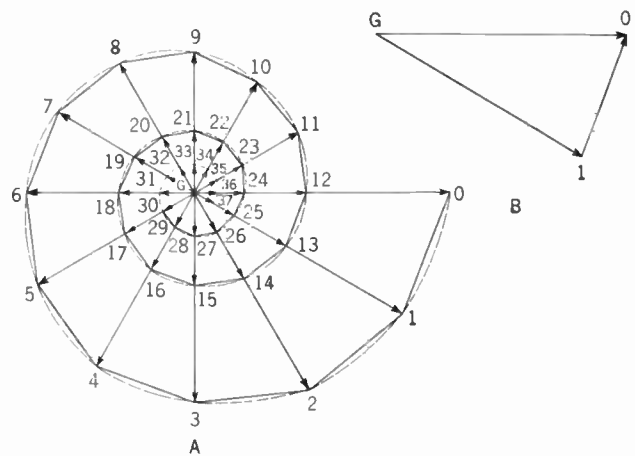

FIG. 15-12 POLAR DIAGRAM OF THE VECTORS OF
FIG. 15-11 (C)

generator, each component of a different magnitude and phase. The effect of these components can be observed, since at certain junctions the line current is flowing in the opposite direction to that taken by the entering current; at other points there is a 90° phase difference between the two; and at still other points there is no phase difference. In other words, the current vector may be considered as moving about G, rotating through 30° for every section traversed and diminishing in value about 10% in each section.

Figures 15-11 and 15-12 show the effective values of the current at certain points along the line and their relative phase positions. These diagrams are independent of time, i.e., they are applicable at any and all times. If on the other hand, we select a given instant of time and plot the instantaneous values of the current at the same points along the line, we obtain the curve shown in Figure 15-13 which shows clearly how the current reverses in direction as it passes through the various sections. It also shows how the current decreases in value or is attenuated as it passes along the line. It will be noted that at the instant shown the input current is at a maximum value flowing in a positive direction; at a point three sections along the line no current is flowing; at a point six sections along the line maximum current is flowing in a direction opposite to that of the current at the input end; at nine sections there is again no current while at 12 sections a maximum positive current is flowing.

The distance between adjacent points where maximum positive (or negative) current is flowing is known as the wavelength and is represented by the symbol λ. The time required to set up one
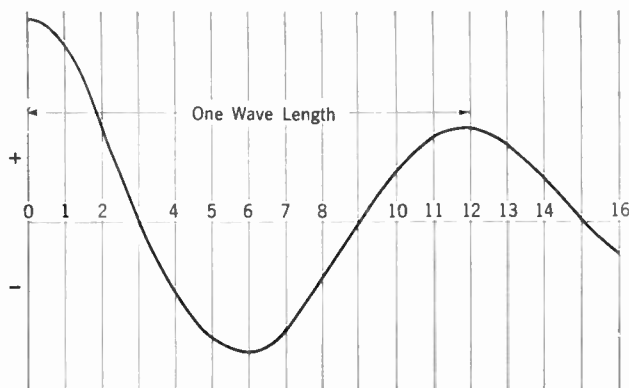
[137]

FIG. 15-13 INSTANTANEOUS CURRENT ALONG UNIFORM TRANSMISSION LINE

wavelength along the line is naturally equal to the time required for the impressed voltage to complete one cycle. The velocity of propagation of the energy along the line is therefore equal to the wavelength divided by the time required to establish the wave, which is the reciprocal of the applied frequency; thus—

$$W = \frac{\lambda}{\frac{1}{f}} = f\lambda \qquad (15:17)$$

This velocity may range in value in telephone circuits from as little as 10,000 miles per second to a maximum approaching the speed of light in a vacuum, some 186,000 miles per second.

The value of the wavelength λ is readily determined from the value of β which, as we have seen, depends on the primary constants of the line and the frequency. β gives in radians the phase shift or delay per unit section. The total phase shift for one wavelength is obviously 360° and since 360° equals 2π radians we may write the equation—

$$\lambda = \frac{2\pi}{\beta} \qquad (15:18)$$

The equation for velocity may then be rewritten

$$W = \frac{2\pi f}{\beta} = \frac{\omega}{\beta} \qquad (15:19)$$

Example: Assuming a 50-mile, 19-gage H-44 side circuit terminated in its characteristic impedance and with an input power at the sending end of 10 milliwatts, calculate at 1000 cycles per second (1) the characteristic impedance, (2) the magnitude of the received current at the distant end and (3) its phase relation with the sent current, (4) the

[138]

power received, (5) the wavelength, and (6) the velocity of propagation. The equivalent distributed constants per loop mile of this particular circuit at 1000 cycles are as follows:

$R = 89.7$ ohms $\qquad C = .062$ μf

$L = .040$ henry $\qquad G = 1.5$ μmhos

Solution:

$$\omega = 2\pi f = 2 \times 3.1416 \times 1000 = 6283$$

From equation (15:8)

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

$$= \sqrt{\frac{89.7 + j6283 \times .040}{1.5 \times 10^{-6} + j6283 \times .062 \times 10^{-6}}}$$

$$= \sqrt{\frac{266.8 \underline{/70°21'}}{389.5 \times 10^{-6} \underline{/89°47'}}}$$

$$= \sqrt{684,980} \underline{/-19°26'}$$

$$= 827.5 \underline{/-9°43'} \text{ Ans. (1)}$$

The input power, P_0, is

$$P_0 = E_0 I_0 \cos\theta$$

Substituting

$$I_0 = \frac{E_0}{Z_0} \text{ in the above,}$$

$$P_0 = \frac{E_0^2 \cos\theta}{Z_0}$$

or

$$E_0^2 = \frac{P_0 Z_0}{\cos\theta}$$

(Note: When Z_0 is a pure resistance, θ is zero and its cosine is one. Therefore, when θ is small in value it may, for all practical purposes, be disregarded.)

$$E_0^2 = \frac{.010 \times 827.5}{.9856} = 8.396$$

$$E_0 = 2.90 \text{ volts}$$

Then

$$I_0 = \frac{E_0}{Z_0}$$

$$= \frac{2.90}{827.5} = .0035 \text{ ampere or 3.5 milliamperes}$$

FIGURE 15-14

[139]

From equation (15:11)

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$$

$$= \sqrt{266.8 \,\underline{/70°21'} \times 389.5 \times 10^{-6} \,\underline{/89°47'}}$$

$$= \sqrt{103,920 \times 10^{-6} \,\underline{/160°8'}}$$

$$= .3222 \,\underline{/80°4'}$$

From equation (15:12)

$$\gamma = \alpha + j\beta$$

$$= .3222 \cos 80°4' + j.3222 \sin 80°4'$$

$$= .0556 + j.3174$$

From equation (15:15)

$$2.303 \log \frac{I_n}{I_0} = -n\alpha$$

or

$$2.303 \log \frac{I_0}{I_n} = n\alpha$$

$$\log \frac{3.50}{I_n} = \frac{50 \times .0556}{2.303} = 1.207$$

$$\frac{3.50}{I_n} = 16.11$$

$$I_n = \frac{3.50}{16.11} = .22 \text{ milliampere} \quad \text{Ans. (2)}$$

Phase shift per mile

$$\beta = .3174 \text{ radian or } 18.2°$$

Total phase shift for 50-mile circuit is

$$50 \times 18.2 = 910° \quad \text{Ans. (3)}$$

Then $I_n = .22 \,\underline{/-910°}$ milliampere

From equation (15:15)

$$2.303 \log \frac{E_0}{E_n} = n\alpha$$

$$\log \frac{2.90}{E_n} = \frac{50 \times .0556}{2.303} = 1.207$$

$$\frac{2.90}{E_n} = 16.11$$

$$E_n = \frac{2.90}{16.11} = .18 \text{ volt}$$

Then

$$P_n = E_n I_n \cos \theta$$

$$= .18 \times .00022 \times .9856$$

$$= .000039 \text{ watt or .039 milliwatt. Ans. (4)}$$

From equation (15:18)

$$\lambda = \frac{2\pi}{\beta}$$

$$= \frac{6.283}{.3174} = 19.79 \text{ miles} \quad \text{Ans. (5)}$$

From equation (15:17)

$$W = f\lambda$$

$$= 1000 \times 19.79$$

$$= 19,790 \text{ miles per second} \quad \text{Ans. (6)}$$

The conditions along the line are graphically illustrated by Figure 15-14. The ordinates of the dashed curves in B and C represent the magnitudes of effective voltage and current at all points throughout the length of the circuit. The voltage and current vectors represent both magnitude and phase relation at the end of each 2-mile section. The instantaneous voltage on the line is shown in D. The power at all points along the line is shown by E. As the power is proportional to the product of $EI \cos \theta$, it decreases faster percentagewise than either the effective voltage or current values illustrated by the dashed curves in B and C.

CHAPTER 16

TELEPHONE TRANSMISSION THEORY—*(Continued)*

16.1 Loading

From the equations for attenuation and velocity developed in the preceding Article it will be seen that in the general case both quantities are functions of frequency. For telephone transmission over long lines, variation of these quantities with frequency is obviously undesirable because it results in distortion of the transmitted signal. Thus frequencies at the upper end of the voice range might suffer more attenuation than frequencies at the lower end of the range; also they might differ from the lower frequencies in the time taken in reaching the receiving point. The seriousness of such distortions in practice of course would depend on the extent of the attenuation and phase shift variations with frequency. Optimum transmission conditions would be obtained with a line where α was as low in value as possible and entirely independent of frequency; and β was so related to frequency that W would be independent of that factor. Such a condition can be realized in theory with a line whose primary constants are of such values that $LG = RC$. When this is true, the value of α as determined from equation (15:13) reduces to—

$$\alpha = \sqrt{RG} \qquad (16:1)$$

and β from equation (15:14) becomes—

$$\beta = \omega\sqrt{LC} \qquad (16:2)$$

Here α is independent of frequency as desired, and β is proportional to frequency in such a way that—

$$W = \frac{\omega}{\beta} = \frac{\omega}{\omega\sqrt{LC}} = \frac{1}{\sqrt{LC}} \qquad (16:3)$$

which is also independent of frequency.

Under the same optimum, if somewhat hypothetical condition, the value of the characteristic impedance works out to be approximately—

$$Z_0 = \sqrt{\frac{L}{C}} \qquad (16:4)$$

Again this is independent of frequency, which is desirable because the characteristic impedance of the line can then be matched by pure resistance.

Unfortunately, a transmission line having such optimum characteristics is not readily attainable in practice. The value of G is actually very small in normal transmission lines and it would not be desirable to increase it artificially because that would tend to increase the attenuation correspondingly. The value of C likewise cannot be changed much because of practical considerations of wire spacing. To attain the optimum condition where $LG = RC$, therefore, it would be necessary either to increase the value of L substantially or to decrease the value of R. The latter is not practical beyond rather narrow limits because of economic considerations. It is possible, however, to increase the value of L to a limited degree and thus approach the conditions sought. This practice is known as loading.

The artificial addition of inductance or loading to a circuit is practicable only to an extent that will approximate the optimum situation discussed above. Reasonably close approaches may be made to the values given for β, W and Z_0 in equations (16:2), (16:3) and (16:4), but the value for α in practice will be more nearly

$$\alpha = \frac{R}{2} \times \sqrt{\frac{C}{L}} \qquad (16:5)$$

which may also be written—

$$\alpha = \sqrt{\frac{R}{2L} \times \frac{RC}{2}} \qquad (16:5)$$

In a cable circuit, which is the only type of circuit to which loading is now applied, the conductance G is negligible and the inductance of the conductors may also be neglected because of their close spacing. If we introduce these assumptions into equation (15:13), then the attenuation constant for the non-loaded cable circuit will approximate—

$$\alpha = \sqrt{\omega\frac{RC}{2}} \qquad (16:6)$$

If this is compared with equation (16:5) it will be seen that the addition of inductance reduces the attenuation to the extent that the value of the

ratio $R/2L$ is less than ω. Since it is generally not feasible in practice to increase inductance without also adding resistance, loading also increases the resistance of the circuit, and this partially offsets the beneficial effect of the added inductance. Nevertheless, material net reduction in the value of α can be obtained. Practical loading also decreases phase distortion as pointed out above; and increases the value of the characteristic impedance, as may be seen by inspection of equation (16:4). This latter gives the effect of transmission at a higher voltage and lower current, which also may be helpful to an understanding of why the attenuation is decreased.

Despite the advantages discussed above, there are important practical limitations to the usefulness of loading in modern telephone practice. One disadvantage is its effect in decreasing the velocity of propagation. Where the loading is "heavy" —i.e., relatively large amounts of inductance are added—the time delay of propagation over very long circuits may become great enough to introduce disturbing effects. Much more important is the fact that practical loading imposes a sharp limitation on the total range of frequencies that can be transmitted. This would not be true if it were feasible to add the required inductance to the circuit on a continuous uniform basis by wrapping each conductor with a spiral of magnetic material such as permalloy. This can be and is done in cases such as long submarine telegraph cable where only a single conductor is involved. But it would obviously be impracticable to apply a similar technique to the many conductors of ordinary telephone cables or open-wire lines. Instead, "lumped" loading is applied by inserting inductance coils at regularly spaced intervals along the lines. This effectively breaks the loaded circuit up into network sections, the major electrical constants of each of which are the series lumped inductance and the shunt capacitance.

Such a network has the essential characteristics of a "low-pass filter" which means that it will tend to block the transmission of frequencies above some critical value. This filtering action is due to resonance effects which are explained in a later Chapter, but it may be noted at this point that the equations given above for the characteristics of a loaded circuit have a definite resemblance to the resonance equation given in Article 13.4. The critical frequency where the attenuation of the loaded circuit begins to increase rapidly is known as the cutoff frequency and may be deter-

mined from the following equation—

$$f_c = \frac{1}{\pi\sqrt{LC}} \qquad (16:7)$$

where L is the inductance of the loading coil in henries and C is the total capacitance of the length of circuit between the adjacent coils in farads.

In the design of loading systems the value of this cutoff frequency may be varied considerably by varying the spacing of the loading points and the amount of inductance inserted. It is not feasible in practice, however, to design a loading system in which the cutoff frequency is much higher than about 30,000 cycles. Loading, therefore, cannot be applied to circuits on which broad-band carrier systems are to be superimposed. Its present application in long distance telephone practice is limited to toll cable circuits on which carrier systems are not superimposed, and to the relatively short toll entrance cables connecting to open-wire facilities.

16.2 Total Losses in the Transmission System

It was pointed out previously that a complete transmission system includes at least three parts —a generating or sending circuit, a transmission line, and a load or receiving circuit. We have seen that there will be attenuation losses on the line in any practical case. Unless the impedance of the sending circuit is exactly equal to the sending end impedance of the line, and the impedance of the receiving circuit is exactly equal to the impedance of the receiving end of the line, there will be other losses known as **reflection losses** at these junction points. Figure 16-1 represents a simple transmission system in which a sending circuit having an impedance Z_S is connected at points 1 and 2 to a line of length l, of characteristic impedance Z_0, and propagation constant γ; which in turn is connected at points 3 and 4 to a receiving circuit of impedance Z_R. Assuming first that the line is so long that both its sending end and receiving end impedance is equal to its characteristic impedance
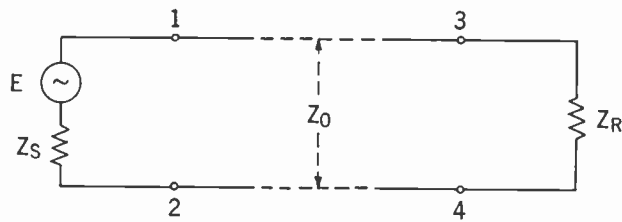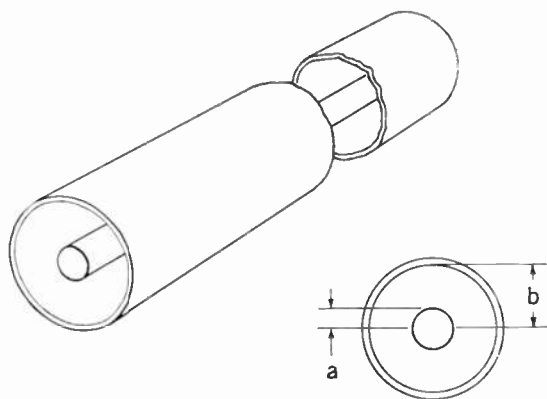

FIGURE 16-1

FIGURE 16-2

Z_0, there will be reflection losses at both junction points if neither Z_s nor Z_R is equal to Z_0. Considering the junction at 1,2 the ratio of the current I_0 entering the line to the current I_0' that would enter the line if Z_s were equal to Z_0 is—

$$\frac{I_0}{I_0'} = \frac{2\sqrt{Z_s Z_0}}{Z_s + Z_0} \qquad (16:8)$$

Similarly the reflection loss at point 3,4 may be determined from—

$$\frac{I_R}{I_R'} = \frac{2\sqrt{Z_0 Z_R}}{Z_0 + Z_R}$$

It may be noted that the criterion for no reflection loss is that the impedances at each side of the junction points should be identical and that this is not the same as the criterion for maximum power transfer, which requires that the impedances be conjugate. The total loss in the transmission system, remembering that an effectively infinite line is assumed, is now equal to the sum of the two reflection losses, determined as above, plus the loss in the line itself, which is $\varepsilon^{-\gamma l}$.

If, as often may be the case, the line is not electrically "infinite" in length, the sending and receiving end impedances will be somewhat different from Z_0 because of the different impedances connected to the distant ends. Any such differences will of course reduce the accuracy of the reflection losses as calculated from the equations given above. An exact general expression for the total loss of the system, however, may be written as follows:

$$\frac{I_R}{I_R'} = \frac{Z_s + Z_R}{2\sqrt{Z_s Z_R}} \times \frac{2\sqrt{Z_0 Z_s}}{Z_0 + Z_s} \times \frac{2\sqrt{Z_0 Z_R}}{Z_0 + Z_R} \times \varepsilon^{-\gamma l}$$

$$\times \frac{(Z_0 + Z_R)(Z_0 + Z_s)}{(Z_0 + Z_R)(Z_0 + Z_s) - (Z_0 - Z_R)(Z_0 - Z_s)\varepsilon^{-2\gamma l}}$$

Here the first three terms represent the two reflection losses as above, and an inverse term correcting for the fact that Z_s is not equal to Z_R; the fourth term $\varepsilon^{-\gamma l}$ represents the line attenuation and phase shift; and the final term represents an "inter-action factor" which takes care of the fact that the line may be so short that the impedance looking into it from either end is not Z_0. Inspection will show that this inter-action factor will approach unity as the length of the line is increased because $\varepsilon^{-2\gamma l}$ becomes very small. Similarly, if Z_s equals Z_R the first term of the equation becomes unity.

The desirability of matching the impedances at each junction point in a transmission system whenever possible is evident.

16.3 Coaxial Conductors

Our consideration of transmission lines thus far has been confined to lines made up of two parallel wire conductors. An entirely different configuration of conductors may be used to advantage where high, and very high, frequencies are involved. This configuration is known as coaxial and the conducting pair consists of a cylindrical tube in which is centered a wire as shown in Figure 16-2. In practice the central wire is held in place quite accurately by insulating material which may take the form of a solid core, discs or beads strung along the axis of the wire or a spirally wrapped string. In such a conducting pair equal and opposite currents will flow in the insulated central wire and the outer tube just as equal and opposite currents flow in the more ordinary parallel wires.

At high frequencies, a unit length of coaxial in which the dielectric loss in the insulation is negligible (effectively gaseous) will have an inductance which is about one-half the inductance of two parallel wires separated by a distance equal to the radius of the coaxial tube. The capacitance of the same coaxial is approximately twice that of two parallel wires separated by the same distance and having the same diameter as that of the central coaxial conductor. If the outside radius of the central conductor is designated a and the internal radius of the tube is b, the characteristic impedance at high frequencies neglecting leakage may be shown to be approximately—

$$Z_0 = \sqrt{\frac{L}{C}} = 138 \log \frac{b}{a} \qquad (16:9)$$

| 143 |

The attenuation constant per mile, where both conductors are of the same material, varies as the square root of frequency and is approximately—

$$\alpha = \frac{R}{2Z_0} = .24 \times 10^{-1} \frac{\sqrt{f}\left(\frac{1}{a} + \frac{1}{b}\right)}{\log \frac{b}{a}} \quad (16:10)$$

where a and b are in centimeters. From equation (16:10) it may be determined that minimum attenuation is obtained when the coaxial is so designed that $b/a = 3.6$. With this configuration Z_0 is about 77 ohms. The present standard co-axial used for transmission in the Bell System employs a copper tube .375 inches in inside diameter and a copper center wire .1004 inches in diameter. This, it will be noted, approximates the optimum ratio specified above for minimum attenuation. The nominal impedance is about 75 ohms. Velocity of propagation in the coaxial approaches closely the speed of light. A study of the basic characteristics of the coaxial shows that at the high frequencies assumed, the attenuation is substantially less than that of a parallel wire line of comparable dimensions. More important is the fact that the shielding effect of the outer cylindrical conductor prevents interference from external sources of electric energy, and also prevents radiation losses of the energy being transmitted over the coaxial.

16.4 Units for the Measurement of Transmission Losses and Gains

As in dealing with any other quantity, we require some unit of measurement when dealing with the energy losses due to attenuation in the transmission of human speech, or in the transmission of any alternating current from a sending device to a receiving device over a long line or through complicated circuits. Without some such unit we would be handicapped in giving any scientific expression to the grade of telephone transmission under various conditions. It would be natural to say that sufficient energy had been transmitted from the speaking station to the listening station for the listener to hear distinctly every spoken word, or to say that the sound coming from the receiver at the receiving station was so faint as not to be intelligible, but this would be a crude method of comparison. For the same reason that we need some adopted standard as a unit of length (such as the foot or the meter) to

measure distance, we require some standard for the measurement of transmission loss or transmission gain in telephone work.

For many years the unit used for this purpose was the **standard cable mile**. This represented the loss due to one mile of an old type of standard 19-gage cable, having a resistance of 88 ohms per mile and a capacitance of .054 μf per mile. In this cable the series inductance and the shunt leakage were negligible, while the bridged capacitance was appreciable. It therefore attenuated the various frequencies that make up the band for telephone transmission unequally, attenuating the higher frequencies more than the lower frequencies. To illustrate, the attenuation constant α was equal to .109 for 800-cycle frequency and .122 for 1000-cycle frequency; etc.

This meant that the percentage reduction in power caused by inserting a mile of standard cable between a sending and receiving element was different for different frequencies. Under these conditions, to say that a telephone circuit had an equivalent of a certain number of miles of standard cable was largely meaningless unless the frequency at which the equivalent was computed or measured was stated at the same time. This rather confusing situation led to the dropping of the mile of standard cable altogether as a unit of measurement and the substitution of a convenient unit not differing greatly in magnitude from the standard cable mile through the voice range, but

TABLE VIII

RELATION BETWEEN DECIBELS AND POWER RATIOS FOR GAINS AND LOSSES

DECIBELS	APPROXIMATE POWER RATIO		
	FOR LOSSES		FOR GAINS
	FRACTIONAL	DECIMAL	DECIMAL
1	4/5	.8	1.25
2	2/3	.63	1.6
3	1/2	.5	2.0
4	2/5	.4	2.5
5	1/3	.32	3.2
6	1/4	.25	4.0
7	1/5	.2	5.0
8	1/6	.16	6.0
9	1/8	.125	8.0
10	1/10	.1	10.0
20	1/100	.01	100.0
30	1/1000	.001	1000.0

having exactly the same significance at any and all frequencies. That is to say, **the new unit, called the decibel (abbreviated "db"), represents always a fixed percentage reduction in power no matter what frequency is involved.** Its magnitude may perhaps be best grasped by remembering that in a circuit equating to ten db the output power will always be one-tenth of the input power. Mathematically, the power ratio for **one** db may be expressed as—

$$\frac{P_1}{P_0} = 10^1 \qquad (16:11)$$

where P_1 is input power and P_0 is output power. This corresponds to a current ratio of $10^{.05}$ and to an attenuation constant value of $x = .115$. Table VIII, showing the power ratios for several values of decibels, will aid in forming a clear conception of the magnitude of the unit.

For any given power ratio the number of decibels corresponding can be determined by the following simple equation—

$$\text{No. of db} = N = 10 \log \frac{P_1}{P_0} \qquad (16:12)$$

or, if the current ratio rather than the power ratio is known—

$$N = 20 \log \frac{I_1}{I_0} \qquad (16:13)$$

or from equation (15:15)—

$$N = 20 \log \frac{I_1}{I_0} = 20 \times \frac{1}{2.303} \times x = 8.68\, x$$

$$(16:14)$$

Although in the above we have been considering the decibel in connection with measurements of loss or attenuation, it is equally useful in the measurement of **gain** such as that given by a telephone repeater. A telephone repeater would be said to have a gain of so many db, for example, if the circuit in which it was used was effectively shortened, or had its net attenuation reduced, to that extent.

17.1 Wave Propagation

In considering the transfer of electric energy along circuits at the lower frequencies, we ordinarily think of a continuous flow of current through the conductors under the pressure of an applied electromotive force. Furthermore, in simple networks or relatively short transmission lines, we ordinarily think of this energy transfer as being practically instantaneous. In the study of long transmission lines and higher frequencies these concepts tend to become inadequate, if not incorrect, as may have already been noted. Thus, even at a frequency as low as 1000 cycles, we would find that the current in a non-loaded 19-gage cable circuit is flowing in opposite directions at points about 88 miles apart, and that an appreciable time is required for the transmission of energy over that distance along the circuit.

Now if we increase the frequency tremendously, we may reach a point where the current is flowing in the line in opposite directions at points only a few feet or even a few inches apart. In other words, the flow of electrons in the conductors is now back and forth within a very limited distance range. Furthermore, because of skin effect this oscillating movement of the electrons is occurring only at, or very nearly at, the surfaces of the conductors. Under these conditions it is somewhat difficult to see any real relationship between the movement of the electrons or current flow and the transfer of energy along the line. It is easier now to think of a **wave of energy** traveling along the line in the space between and surrounding the conductors, and setting up "ripples" of electron motion in the wires as it goes along.

This traveling wave of energy is usually thought of as an electromagnetic field which, in the present case, is associated with or bounded by the wire conductors but which can also exist in free space where no conductors are present. Under this concept the electromagnetic field is considered as being made up of (1) the electric field whose lines of force radiate at right angles from the charges on the surfaces of the conductors and (2) the magnetic field whose lines of force encircle the conductors and are always at right

angles to the force lines of the electric field. The total energy of the wave is always equally divided between the electric and magnetic fields and passes back and forth between them at a rate determined by the frequency of the applied emf. The entire electromagnetic wave travels along the line at the speed of light if not retarded by inductance or capacitance effects in the conductors.

17.2 Reflections on Transmission Lines

If a transmission line were actually of infinite length and without loss, an electromagnetic wave of energy would obviously travel along the line forever; and this would be true even though the sending-end source of energy was disconnected some time after the wave was started. Practically, of course, we are concerned with delivering power to receiving devices over lines of finite length. It becomes necessary, therefore, to consider what will happen when a traveling wave reaches the end of a uniform line.

For purposes of such analysis, it is convenient to assume a line of such low resistance as to be practically without loss, a situation which is actually approximated in a relatively short line at very high frequencies. Let us investigate first the situation where such a lossless line is open at the far end. In that case, there can be no current flow at the open end and the magnetic field therefore

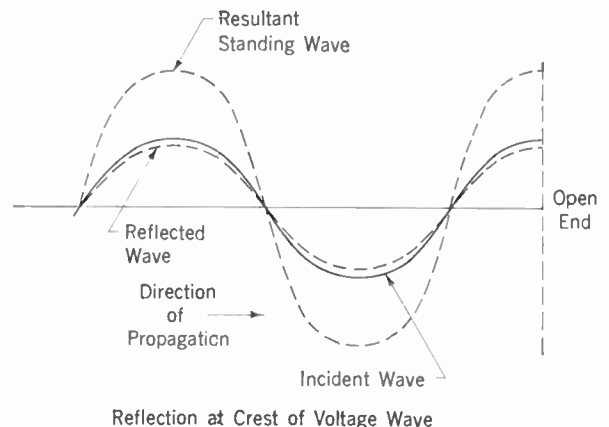Reflection at Crest of Voltage Wave

FIGURE 17-1

must disappear. But the energy which the magnetic field contained must be conserved by being transferred to the electric field. This energy, added to the energy already contained in the electric field approaching the end of the line, doubles the total electric field at that point. In other words, the voltage at the open end of the lossless line becomes twice as great as the voltage applied at the sending end.

We may then consider that the increased voltage starts a wave of energy back along the line towards the sending end. Thus the traveling electromagnetic wave reaching this total discontinuity in the line is **reflected** in somewhat the same way as a light wave is reflected from a mirror. If the initial or incident wave is a sine wave the reflected wave must also be a sine wave. Back over the line, the reflected wave will add to the incident wave at all points and the sum must also be a sine wave. This is graphically illustrated in Figure 17-1 where an incident voltage wave is shown reflected from the open end of the line at its positive crest. The reflected wave then adds in phase with the incident wave to give a wave having crest values approximately twice those of the incident wave. The resultant wave is known as a **standing wave** because it does not travel longitudinally along the line. In other words, if a voltmeter is used to make measurements along the line it would read zero at the nodal points and maximum values of twice the effective value of the applied voltage at the points where the positive and negative voltage crests are shown. Instantaneously at these points the voltage would be varying from maximum positive value through zero to maximum negative value in accordance with the frequency of the applied voltage. This is perhaps better illustrated by Figure 17-2.

A reflection from an open-end line that results in the development of a standing voltage wave of approximately double magnitude will occur only when the distance from the sending end to the open end is equal to an odd number of quarter wavelengths. The illustrations given, for example, show a line one and a quarter wavelengths long. It may also be noted that if, as we have assumed, the line is lossless the impedance at the sending end in this case is theoretically zero. Energy is nevertheless surging back and forth in the line. The line may accordingly be thought of as behaving like a series resonant circuit. For this reason, lines having discontinuities which cause reflections and consequent standing waves are
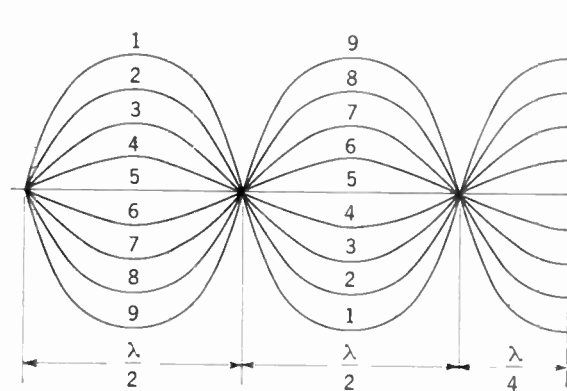

FIGURE 17-2

known as **resonant** lines. A uniform line of infinite length or one terminated in its characteristic impedance, on the other hand, is a non-resonant line. Of course the complete no-loss condition we have assumed for discussion purposes never actually exists, but at very high frequencies the inductive and capacitive reactances may be so large as compared with the resistance as to cause the actual situation to approach closely the no-loss condition.

A similar analysis of a line consisting of an odd number of quarter wavelengths and shorted at its distant end will show that the standing wave of voltage will have its maximum value equal to the maximum value of the applied voltage at the sending end and will have a nodal point or zero voltage at the shorted end. The sending end impedance of such a line approaches infinity. Lossless lines whose length is an even multiple of quarter wavelengths, on the contrary, have extremely high impedance when open at the distant end and impedance near zero when shorted at the distant end.

The extreme situations that we have discussed above would not occur in a transmission line designed to carry energy from one point to another. We have considered only lines that were open or shorted, and only at points that were some integral multiple of a quarter-wavelength at some specific frequency. As will be brought out in a later Chapter, short resonant lines of these types have numerous useful applications in very high frequency work, but they are not useful as transmission lines in the ordinary sense. Nevertheless, the phenomena of wave reflection must be taken into consideration in any transmission line where the impedances are not matched at all junction points. Where there is any such impedance mis-

match on any type of transmission line, and regardless of the frequencies being transmitted, there will be some reflections. In general, the impedance irregularity will not only prevent a maximum power transfer at the point of mismatch but will also result in actual loss of energy. This is due to the fact that any actual transmission line must have some resistance which will cause I^2R losses due to the reflected currents in addition to the I^2R losses of the useful current transmitted to the load at the receiving end of the line. The extent of a reflection loss of course depends on the extent of the mismatch and can be calculated from the equation given in Article 16.2.

17.3 Electromagnetic Radiation

In analyzing the total energy losses that may occur in the transmission of electric power from one point to another, it is necessary now to consider another phenomenon which has, up to this point, been ignored. This is the loss due to **radiation.**

In 1864, James Clerk Maxwell undertook to set up a series of mathematical equations that would provide a general statement of the relationships between electric and magnetic fields under any and all conditions. In rounding out this series of equations to achieve mathematical symmetry, he was led to some very interesting conclusions. The equations seemed to indicate that the hitherto existing assumption that all of the energy contained in the electric and magnetic fields accompanying the flow of current in a conductor returned into the conductor when the source of emf was cut off was not wholly true. Some part of the field, it appeared, would detach itself entirely and escape into space in the form of electromagnetic radiation, carrying with it a comparable part of the total energy.

This led to the further conclusion that a moving electric field can exist in the absence of any electric charges, despite the fact that an electric field is usually thought of as being made up of lines of electric force always terminating on electric charges. Furthermore, if a moving electric field can exist independently in space, it must be thought of as being equivalent in certain ways to a flow of electric current. In other words, the moving electric field in free space must set up, or be accompanied by, a moving magnetic field just as it would be in the case of current flow along conductors.

The Maxwell equations cannot be written in any form that does not involve branches of mathematics which are beyond the scope of this book. With respect to radiated energy, however, the equations indicate, and experience confirms, that for a given current the amount of energy radiated depends upon the square of the frequency. Naturally the amount of radiation also depends upon the intensity of the current. What is of major significance from a practical viewpoint is the fact that the amount of radiation goes up very rapidly as the frequency increases. There is always radiation whenever there are changing current values but at voice frequencies, and at frequencies well up into the ordinary telephone "carrier range", the amount of radiation is negligible for most practical purposes. At frequencies that are measured in megacycles, on the other hand, radiation may cause losses that are much greater than any I^2R losses in the conductor.

17.4 Radio Transmission

It was not until some twenty years after Maxwell developed his famous equations that Heinrich Hertz demonstrated experimentally the truth of the electromagnetic radiation hypothesis; and it was some years later before experimenters began to develop methods for taking advantage of this radiation phenomenon to transmit electric energy through space for useful purposes. In this case the objective was not to avoid energy losses by radiation but to do everything possible to facilitate maximum radiation. Since, as we have seen, the amount of radiation increases with frequency at a geometric rate, purposeful radio transmission naturally involves the use of high frequencies and transmission lines or antennas designed to radiate maximum energy.

If the transmission of the radiated energy were through unobstructed space in the form of electromagnetic waves like light waves, there would be no loss of energy "along the line" because there would be nothing to absorb the energy. If energy could be radiated from a given point and confined in a narrow beam extending directly to the receiving point, this means of transmission could be far superior to any wire transmission because of this lossless quality. However, the natural tendency of any radiator, isolated in space, is to send out energy in practically all directions although only such energy as actually reaches any receiving point is useful.

[148]

"Line losses" in radio transmission should largely be thought of not as energy losses in the transmission path itself, such as occur in wire lines, but as energy escaping entirely from the effective transmission path. The basic transmission problem, accordingly, is to devise methods that will direct the path of the radiated energy. Such methods are concerned primarily with the radiating antennas and it is customary in radio work to measure the effectiveness of antennas in controlling the directivity of radio propagation in terms of **antenna gain.** This is merely a measure in decibels, or other appropriate units, of the amount of energy received at a given point from a given transmitting antenna compared with what would have been received if the transmitting antenna radiated with uniform strength in all directions. Thus, high antenna gain in radio transmission corresponds to low line loss in wire transmission.

As is discussed briefly in a later Chapter, it is possible to design antennas with quite high gains —particularly in the superhigh-frequency range (thousands of megacycles). There is no practical possibility, however, of designing antennas with such directivity that all, or even a major part, of the transmitted energy will reach a receiver located at any great distance from the transmitter. In other words, there must always be a very substantial effective loss of energy. What is perhaps worse, this lost energy may be absorbed elsewhere where it may interfere with other communication circuits, or be received at unauthorized points in such a way as to militate against the privacy of the transmission. There remain, therefore, obvious advantages in the employment of physical facilities that guide the energy directly to the desired receiving point. An ordinary wire line is one type of such a guide but is satisfactory at only relatively low frequencies because its radiation and other losses become too great at high frequencies. A coaxial is a better guide because the outer tube acts as a shield to prevent any of the electromagnetic energy transmitted within the tube from radiating into space. At frequencies in the thousands of megacycles, however, the losses of any practical design of coaxial also become very high.

The usefulness of radio transmission in practice is found in two principal situations. The first is where it is impossible or economically impractical to construct physical facilities that will guide the energy. Here the application is in transmission over large bodies of water or to moving points such as ships at sea, motor vehicles, trains, and aircraft. The second situation is point-to-point service over land where radio transmission is considered more economical on an overall basis than other methods. A major application here is to microwave radio relay systems. But even in this case, it is necessary to use some form of physical transmission medium to guide the energy through the relatively short distances from the transmitter itself to the transmitting antenna, and from the receiving antenna to the receiver. Within limits, shielded wire lines or coaxials can be used for this purpose. More effective, however, are the simple hollow metal tubes commonly known as **waveguides.**

17.5 Waveguides

It is convenient to think of a waveguide as a device which merely isolates a particular radio frequency transmission path in space. Propagation of electrical energy through the guide may then be considered as essentially no different from ordinary radio propagation, except that it is wholly confined to this particular isolated path. However, the cross-sectional dimensions of the guide must be sufficiently large to permit the traveling electric and magnetic fields to assume configurations comparable to those that they would naturally assume when traveling in free space. This means, in mathematical terms, that the field configurations in any given guide, and in a given frequency range, must be such as to satisfy Maxwell's equations. Broadly speaking, it may be said that for effective propagation a waveguide must have a cross-sectional dimension that is in the order of magnitude of at least one half wavelength of the transmitted energy. This limits the practical use of waveguides to the transmission of very high frequencies where wavelengths are no greater than a few inches.

Within the broad limits outlined above, there are a number of possible field configurations that will permit the transmission of energy through a given waveguide. These possible configurations, known as "modes", can be classified into two major categories. In one category are the "transverse electric", or TE, modes in which the electric field is always transverse to the axis of the guide, while the magnetic field has components along the guide axis. The second category includes the reverse situation where the magnetic field is always
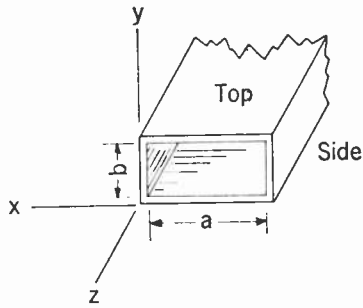
FIG. 17-3 RECTANGULAR WAVEGUIDE

transverse to the axis and components of the associated electric field are in the axial direction. These are known as "transverse magnetic" or TM, modes. Within each of these major categories, a number of different modes are usually possible. These are ordinarily designated by a double numerical subscript such, for example, as TE_{10} or TM_{11}. In a guide of rectangular cross-section, the first numerical subscript indicates the number of half-cycle variations of the field in the direction of the long side of the rectangle; the second subscript gives the same thing in the direction of the short side.

Propagation of a particular mode in a particular guide is possible only when the wavelength of the applied energy has less than a critical value known as the "cutoff wavelength". Any lower frequency is effectively rejected by the guide. The same guide, however, may permit propagation in some other mode because different modes have different cutoff values. For a given design of guide, the particular mode that has the highest cutoff value is designated the "dominant" mode. This is ordinarily the mode most desirable for waveguide transmission since it permits the use of the minimum sized guide for transmission of a given frequency.

Many of the waveguides employed in microwave radio relay systems are rectangular in cross-section as illustrated in Figure 17-3. While the ratio of the two dimensions is not extremely critical, it is usually two to one, as indicated here. The dominant mode in this design of guide is TE_{10}, which means that the preferred field configurations are as shown in Figure 17-4. The electric field is transverse to the axis of the guide, with its lines extending vertically between the top and bottom walls. As indicated, the intensity of the field is maximum at the center and tapers off sinusoidally to zero values at the side walls. The

magnetic field, as always, is at right angles to the electric field so that its lines form closed loops in planes parallel to the top and bottom walls and extending between and along the side walls as shown in Figure 17-4(B). The magnetic field is thus uniform between the top and bottom walls and has its maximum intensity along the side walls and minimum in the center of the guide, where the electric field is strongest.

The cutoff wavelength of the dominant mode of a waveguide proportioned as in Figure 17-3 is equal to $2a$, twice the length of the longer side. The next higher mode, TE_{20}, is cut off at a wavelength equal to a, while still higher modes have still smaller cutoff wavelengths. This means that the band of frequencies whose wavelengths lie between a and $2a$ are propagated in this waveguide in the dominant mode only. All other modes are automatically suppressd. This greatly simplifies various waveguide transmission problems since it permits working with only one, known set of field configurations for the transmission of a considerable band of frequencies.

Although energy losses in rectangular waveguide propagation are kept down by confining the field configurations to the dominant mode, attenuation in the guide is nevertheless high. This is due to the currents induced in the inner walls of the guide by the traveling waves. The existence of such currents may be predicted from consideration of the effect of either the magnetic or the electric fields. Thus, referring to Figure 17-4(B), it will be noted that the lines of magnetic flux extend both axially along the guide and transversely across its longer cross-sectional dimension. As this flux varies in strength, it tends to induce cur-
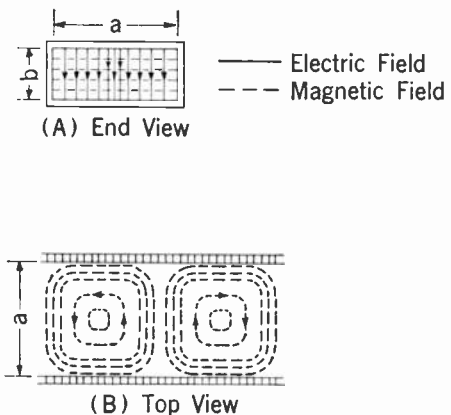

FIG. 17-4 TE_{10} FIELDS IN RECTANGULAR WAVEGUIDE

rents in any adjacent conducting material, flowing at right angles to the direction of the flux lines. The result in this case may therefore be expected to be instantaneous currents flowing vertically in the side walls due to the relatively strong axial magnetic flux adjacent to these walls. The same axial flux also induces currents in the top and bottom walls, flowing transversely between the center line of these walls and the sides. In addition, the transverse magnetic flux may be expected to induce axial current flow in the top and bottom walls, which will be of maximum strength at the center line. Consideration of the electric fields, as illustrated in Figure 17-4(A) leads to the same result. The transverse lines of the electric field extending vertically between the top and bottom walls establish a potential difference between these walls, which causes current flow vertically in the side walls. Further, since the electric field is stronger at the center than at the side walls, transverse currents will be set up between the center line of the top and bottom walls and the side walls. Finally, the sinusoidal shape of the traveling wave, produces peaks and valleys in the moving electrical field, thus creating instantaneous potential differences axially along the guide, which in turn produce axial currents strongest along the center lines of the top and bottom walls.

In considering waveguide transmission, it is convenient to think of these currents in the walls as in the nature of eddy currents which merely absorb energy from the traveling waves as a consequence of I^2R losses. Their attenuating effect, however, is considerable. The loss at 4000 megacycles of a $1\frac{1}{4} \times 2\frac{1}{2}$ inch bronze guide is about 1.5 db per 100 feet. This is extremely high in comparison with the losses of ordinary wire transmission lines at low frequencies. It is nevertheless substantially lower in the microwave region than would be caused by the usual type of wire line or coaxial cable. The velocity of propagation of energy in a waveguide approaches but is always somewhat less than the speed of light.

As in any other type of transmission line, it is desirable that a waveguide be uniform in structure along its total length. Any discontinuity such as changes in size or shape, holes in the guide walls, or foreign conducting materials in its interior, will distort the configuration of the traveling electric and magnetic fields to a greater or less extent. This distortion will, in general, cause the generation of new and unwanted modes which reduce the energy of the applied dominant trans-

mission mode. These new modes may or may not be able to propagate in the guide. If so, they will usually contribute nothing to the net effective transmission. If not, their effect will be like that of a purely reactive load, causing some portion of the dominant wave to be reflected back toward the source with consequent additional losses. Reflection losses may also be caused by too sharp bends or twists in a waveguide, as well as by improper terminations at either the sending or receiving end. When irregularities cannot be avoided, various methods of "impedance matching", similar in principle to those employed in wire or coaxial lines, are used to avoid or at least minimize reflection losses. Some of the methods and devices employed for this purpose are discussed in succeeding Chapters.

In addition to avoiding discontinuities that might result in the generation of unwanted modes, it is of course necessary to energize the guide in such a manner that only the desired dominant mode is produced in the first place. In the case of the rectangular guide that we have been discussing, this can be accomplished almost ideally by the method shown in Figure 17-5. Here an energized coaxial probe is connected to the waveguide midway between its sides and at such a distance from the shorted end that reflected energy will return in phase with the forward transmitted wave. Current flowing in the center coaxial conductor will set up a concentric magnetic field which will lie in planes parallel to the top and bottom of the guide. The IR drop in the probe also establishes a potential difference between the top and bottom of the guide which results in a transverse electric field extending vertically between top and bottom, with its maximum intensity at the center. Thus, the desired dominant mode, coded TE_{10}, is automatically produced.

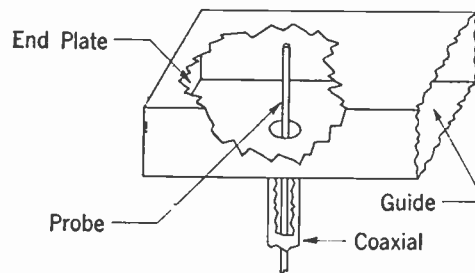Wave propagation in guides of circular cross-section requires the use of somewhat different

FIG. 17-5 PROBE TYPE COAXIAL TO WAVEGUIDE TRANSDUCER

techniques than in the case of the rectangular guides. The dominant mode in a circular waveguide, coded TE$_{11}$, is illustrated in Figure 17-6. It will be noted that this is similar to the TE$_{10}$ dominant mode of the rectangular guide. Its cut-off wavelength is closer to that of the next higher mode, however, than is the case in the rectangular guide, which limits the frequency range over which operation in a single mode can be assured. In most practical applications, accordingly, possible propagation of a number of modes must be expected in the circular guide and the use of mode filters or other means to suppress the unwanted modes is generally necessary. On the other hand, at frequencies in the order of 10,000 mc attenuation in the circular guide is appreciably lower than in a rectangular guide of comparable size. This makes the circular guide more attractive for use when it is desired to transmit several bands of frequencies through a single guide as in TD, TH and TJ microwave systems where the signal frequency bands center about 4000, 6000 and 11000 mc, respectively.

In circular waveguides, it is necessary also to consider the polarization of the transverse electric field. Thus, in Figure 17-6 the electric field lines are shown as extending vertically across the guide —as they are also in the rectangular guide of Figure 17-4. The geometry of the rectangular guide does not permit the lines to take any other than this configuration for the mode illustrated. The circular guide, however, is perfectly symmetrical in all transverse directions and thus provides no positive means of holding the field lines in the vertical position. In other words, there is nothing to prevent the field from rotating into a horizontal, or any other, position as the wave travels along the guide. Such behavior could seriously complicate methods for removing energy at the receiving end of the guide, and otherwise controlling its transmission. Means are available, however, which permit precise control of the wave
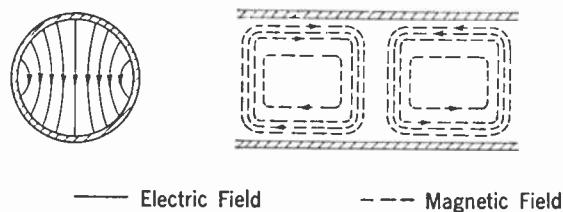
polarization. These depend upon certain interesting properties of magnetic ferrites and are discussed briefly in a later Article.

Although it has not yet been applied in telephone practice, it is hardly possible to conclude this discussion of waveguide transmission without mention of the circular electric, or TE$_{01}$, mode in the circular guide. This mode, which is illustrated in Figure 17-7, has the completely unique property of continuously decreasing attenuation loss as frequency is increased. It thus offers a real possibility for waveguide transmission of very broad signal bands over practically unlimited distances. Theory and experiment indicate that a two inch diameter copper pipe operated in this mode at frequencies in the order of 50,000 mc could transmit a band of frequencies at least 500 mc wide with a loss no greater than about 2 db per mile. This compares favorably with the losses encountered in current microwave radio relay practice.

The reason for the low losses in this mode may be understood by referring again to Figure 17-7. Here the lines of electric force do not extend to the walls of the guide as in all other transverse electric modes, but close upon themselves. They accordingly set up no charge accumulations at half-wave intervals along the guide walls to produce the longitudinal flow of current that is the principal cause of attenuation in other modes. There is some circular current flow in the guide walls but this is limited to that necessary to produce only such counter magnetic field as is necessary to prevent the magnetic fields within the guide from escaping through the walls. Furthermore, as the frequency increases the magnetic fields within the guide tend to contract toward the center so that, in theory at least, the attenuation would become zero at infinite frequency.

As can usually be expected, practical attainment of the results indicated by the above theory presents some serious problems. Establishment of the TE$_{01}$ mode in the guide is not too easy in the first place. Maintenance of the mode in full purity as the wave travels along the guide is practically impossible. Any mechanical irregularity in the walls of a solid tubular guide, including the slightest deviation from straightness, will result in the conversion of some part of the energy into one or more of the many other modes of propagation that are possible. This not only causes some increase in attenuation losses but may produce more serious distortion effects when the
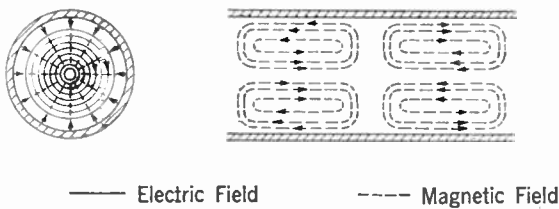

——— Electric Field   – – – Magnetic Field

Fig. 17-6 TE$_{11}$ Fields in Circular Waveguide

——— Electric Field   ----- Magnetic Field

FIG. 17-7 TE₀₁ FIELDS IN CIRCULAR WAVEGUIDE

guide is of considerable length. This is due to the fact that some of the energy of the major mode may be converted into a different mode at one point and then reconverted back to the original mode at some other point of irregularity farther along the guide. In most cases, the velocity of propagation of the two modes will differ. The net result may then be that the reconverted energy is no longer in phase with that of the major transmission. The magnitude of signal distortion produced will of course depend on the amount of mode converted energy involved and the distance between the points of conversion and reconversion.

It would naturally be expected that waveguides designed for long distance transmission would be built with at least as much care as the relatively very short guides now in use. Complete elimination of irregularities, including bends, is, however, obviously impractical. It follows that successful application of really long guides will require the employment of methods that effectively suppress unwanted modes. A number of possible methods have been suggested. Most promising is a circular guide structure consisting of a continuous insulated copper conductor wound into a helix of very small pitch and supported within a tube of "lossy" material. This is effectively equivalent to a structure made up of a series of individual copper rings insulated from each other and likewise surrounded on the outside by a lossy housing. Since the conducting rings are perpendicular to the axis of the guide, such a structure readily supports the TE$_{01}$ mode as well as the higher order circular electric modes. It cannot support propagation of any other TE modes, or of any TM mode, however, because no appreciable path exists for the longitudinal guide wall currents that must be present in all but the circular electric modes. The possible propagation of the higher circular electric modes such as TE$_{02}$ or TE$_{03}$ does not present a serious problem because experiment has shown that substantial conversion of the TE$_{01}$ mode to these higher modes of the same family is not likely to occur.

Somewhat similar results can be obtained by coating the inner wall of a solid copper tube with a thin layer of insulating material. This likewise tends to suppress all but the circular electric modes, while adding little to the loss encountered by the desired TE$_{01}$ mode.

17.6 Cavity Resonators

If one end of a waveguide is "shorted" (i.e., closed by a conducting end plate) total reflection of any applied energy will occur just as when a wire transmission line is shorted. And, as in the case of the wire line discussed in Article 17.2, if the short is at a point which is an odd multiple of a quarter wavelength from the energy source, the reflected wave will add in phase to the incident wave to set up a standing wave in the guide. In other words, the guide now behaves as a resonant line with relatively large amounts of energy surging back and forth between its contained electric and magnetic fields. This phenomenon is taken advantage of to produce the cavity resonators that we shall encounter in later Chapters. Thus, if a section of waveguide one half wavelength long is closed at both ends to form a small box and energized at the center point, it may be considered as two quarter wave lines, each of which is resonant. The cavity then acts like a parallel resonant circuit made up of an inductor and a capacitor where the application of relatively small exciting energy may produce very large energy oscillations at the resonant frequency.

As indicated above, the frequency of resonance of any given cavity is determined by its internal dimensions in terms of wavelengths. It follows that practically any enclosed metallic (i.e., conducting) structure of whatever size or shape may behave as a resonator at some frequency. Moreover, as in waveguides, the existence of many different field configurations, or modes, is possible within any given cavity. The resonant frequency would of course be different for each such mode. The particular mode which permits resonance at the lowest frequency is designated the dominant mode and this is the only mode with which we need be concerned in most practical applications of cavity resonators.

17.7 Waveguide Isolators

A device of great usefulness in waveguide
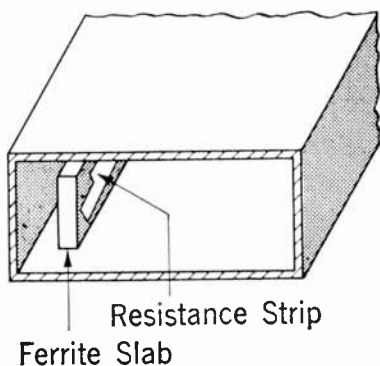
Resistance Strip
Ferrite Slab

FIG. 17-8 FIELD DISPLACEMENT ISOLATOR

transmission is the so-called isolator. This is a waveguide element which effectively permits transmission through a guide in one direction only. It can thus be used to prevent energy, reflected from a transmitting antenna or other discontinuity, from traveling backward through the guide to points where it might adversely effect the operation of vital circuit components such as klystron oscillators or traveling-wave tube amplifiers.

The non-reciprocal behavior of isolators depends upon certain peculiar properties of the magnetic ferrites whose use in high-frequency transformer cores was discussed briefly in Article 14.6. Because of their extremely high resistivity, ferrite materials do not normally have any appreciable effect on the propagation of high frequency electromagnetic waves. If inserted within a waveguide, they are, in other words, practically transparent to a microwave traveling through the guide. In the presence of an externally applied d-c magnetic field, however, the behavior of the ferrite is quite different. Within the ferrite, because of its unique magnetic structure, there is now definite interaction between the applied field and the magnetic field of the traveling microwave. This effectively changes the permeability of the ferrite, which results in some distortion or displacement of the electric field of the traveling microwave. Moreover, the interactions are such that the permeability change in the ferrite is different for the two directions of microwave transmission. The electric field displacements are accordingly also different.

Figure 17-8 shows a section of rectangular waveguide in which a thin slab of ferrite, which might be typically a few inches in length, is placed in an off-center position. A d-c magnetic field is

assumed applied to the ferrite in a direction at right angles to the magnetic field of the traveling microwave. Such an arrangement will displace the electric fields of forward and reverse traveling microwaves in some such manner as indicated in Figure 17-9. This shows at A the normal electric field intensity distribution in a cross-section of rectangular waveguide; and at B, the way in which this field is displaced by the magnetized ferrite for both forward and reverse waves. It may be noted that the field intensity of the forward wave is minimum at the ferrite slab while the intensity of the reverse wave is maximum at the same point. Now, if a coating of resistive material such as graphite is applied to one side of the slab, as shown in Figure 17-8, it will absorb most of the energy of the reverse wave. The forward wave, on the other hand, will be attenuated only slightly by the resistive coating because of its low field intensity at that point. The magnitude of the discrimination between forward and reverse waves depends rather critically on such factors as the length and width of the ferrite slab, its location with respect to the guide wall, the strength of the applied magnetic field, and particularly the length, shape and exact location of the resistive strip. In practice, representative losses introduced by field displacement isolators of this type are about .2 db in the forward direction of transmission and 30-35 db in the reverse direction.

In circular waveguide, advantage may be taken of another special property of magnetized ferrite to effect microwave isolation. When a short rod
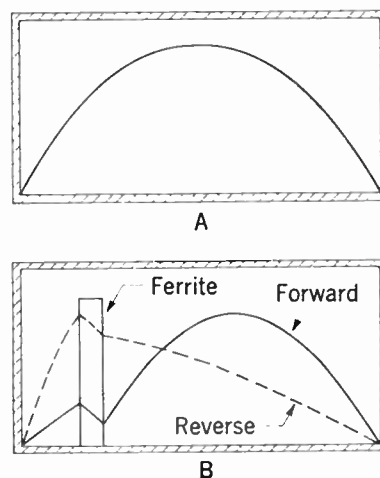
FIG. 17-9 ELECTRIC FIELDS IN RECTANGULAR
WAVEGUIDE

or pencil of ferrite material is centered in a circular waveguide and magnetized in a longitudinal direction by an external magnet, it is found that the plane of polarization of a transverse electric microwave in the guide is rotated to a new angular position as it passes the ferrite. The extent of the rotation is accurately determined by the size and shape of the ferrite pencil and the strength of the applied magnetic field. Of particular interest, moreover, is the fact that the field rotation is in the same angular direction regardless of the direction in which the microwave is traveling through the guide. This rather remarkable behavior is called "Faraday rotation" because of its close analogy to a comparable optical phenomenon first noted by Michael Faraday more than a century ago.

How this Faraday rotation effect is employed in microwave practice to secure isolation may be understood by reference to Figure 17-10. Here a thin pencil of ferrite, tapered at both ends to minimize reflections, is held centered in a section of circular waveguide. An axial magnetic field is established in the ferrite pencil by means of a permanent magnet surrounding the section of waveguide, as shown. It is assumed here that the dimensions of the ferrite and the strength and direction of the magnetic field are such that the polarization of a wave in the guide will be rotated exactly 45° in a counter-clockwise direction, as viewed from the left end of the guide. At each end, the waveguide is transformed from circular to rectangular cross-section by a gradual transition that prevents impedance irregularity. The transition at the right end, however, is so made that the rectangular guide there is turned through 45° with respect to the left end rectangular guide.
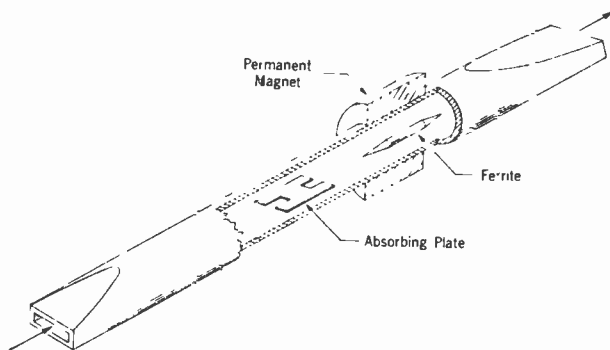
From the general discussion of waveguide transmission in Article 17.5, it will be recalled that for normal TE_{10} transmission in a rectangular guide, the electric field lines always extend between the longer walls and parallel to the shorter walls. In Figure 17-10, this configuration may be designated as vertical polarization of the electric field. In a circular waveguide, on the other hand, the dominant mode TE_{11} may be transmitted equally well in any plane of polarization. If the desired direction of transmission in Figure 17-10 is from left to right, a signal entering the guide at the rectangular left end will necessarily be vertically polarized. As it passes through the transition into the circular guide, it will continue to be polarized vertically. Passage through the ferrite, however, will rotate the polarization 45° counter-clockwise. But since the right rectangular guide is also turned through 45°, the rotated wave will pass into the rectangular guide at the right still in the vertical plane with respect to the guide. Transmission is thus effected without appreciable attenuation of the wave.

But now if the transmitted wave is partially or completely reflected by some irregularity farther along in its path, the reflected wave will encounter a different set of conditions. It will return still vertically polarized until it reaches the ferrite. Here it will be again rotated through 45° but also in a counter-clockwise direction as viewed from the sending end. Its plane of polarization is now turned exactly 90° from that of the original input wave. The rectangular guide cannot transmit a wave in this horizontal polarization and the wave must therefore be totally reflected when it reaches the transition. If desirable, this second reflection can be largely avoided by inserting a plate of energy absorbing material between the ferrite and the transition, as shown in the drawing. This may consist of a small card of insulating material which has been given a thin coating of a "lossy" substance such as carbon or aquadag that readily soaks up microwave energy. Mounted in a horizontal plane, as indicated, the card effects high attenuation of the reflected wave while offering practically no opposition to the transmission of the vertically polarized forward wave.

In practice, it may not always be feasible to control the rotation of the wave polarization to the precise angle that is required for 100% isolation. A simple device such as that here pictured, however, may be expected to introduce a loss of only about .25 db in the desired direction of trans-

Permanent Magnet

Ferrite
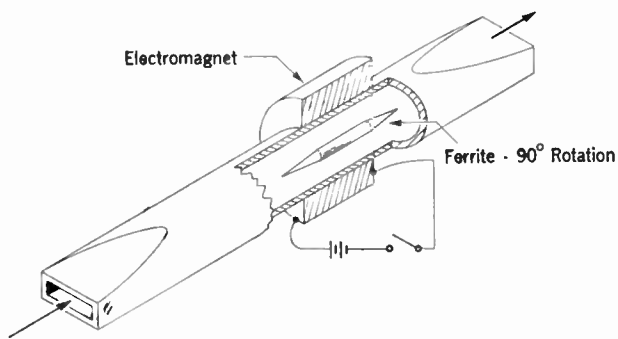
Absorbing Plate

FIG. 17-10 FARADAY ROTATION ISOLATOR

FIG. 17-11 FARADAY ROTATION WAVEGUIDE SWITCH

mission, and from 30 to 35 db in the opposite direction. When more complete isolation is desired, several arrangements employing two or more ferrite units in tandem have been designed.

The Faraday rotation effect is also used to produce the high-speed waveguide switches used in certain microwave radio systems. Here the wave rotation is controlled by an electro-magnet rather than a permanent magnet and the magnetization is of such value as to rotate the wave through 90°. When a section of circular guide, so equipped, is inserted in a rectangular guide as shown in Figure 17-11, it has no appreciable effect on transmission as long as the electro-magnet is not energized. Simply closing the circuit through the electro-magnet results in complete blocking of transmission through the guide. Thus, two such devices make possible almost instantaneous switching of the wave transmission from one path to another.

CHAPTER 18

CHARACTERISTICS OF CIRCUIT FACILITIES

18.1 Classification of Wire Facilities

Transmission facilities or media employed in telephone and telegraph work, to be cataloged completely, would have to include both waveguides and free space, which is the medium of radio transmission. In this Chapter, however, we shall confine our attention to various types of wire conductors, including coaxials. The kind of wire line facility to be used in a particular case depends upon economic considerations and the transmission requirements to be met. Ordinary wire facilities may be classified in several ways according to their uses, or on the basis of their physical or electrical characteristics.

It is customary first to make a general distinction between facilities used for toll (long distance) and for exchange area transmission. The latter facilities include the greater part of the total telephone plant since local or short haul service is naturally used more frequently than long distance service. Accordingly, it is economically desirable to design these facilities primarily on the basis of providing satisfactory transmission within the exchange area. For toll or long distance connections, of which local facilities necessarily form a part in every case, more costly types of facilities are used for the long distance links in order that the transmission shall remain satisfactory. This arrangement is in the interest of overall economy because the long distance facilities are relatively few as compared with the local facilities. It means in general that the latter facilities do not have to meet as exacting requirements as do the toll facilities with respect to attenuation per unit length, impedance regularity, or balance against noise and crosstalk. In exchange area cables, for example, wire conductors as fine as 22, 24, or 26-gage are widely used, whereas the minimum gage in long toll cables is 19. Generally similar distinctions as between local and toll transmission apply in the case of open wire facilities. However, it may be noted that there is a certain middle ground where exchange area trunks are of such great length in some cases that their transmission requirements are not widely different from those of the shorter toll circuits. Loading is frequently applied to such trunks and in some cases it may be necessary to use telephone repeaters as well.

The principal types of toll or long distance wire facilities are considered separately in the following Articles.

18.2 Open Wire Facilities

In both open wire and cable circuits, the development of the telephone art has involved the use of many different types of circuit facilities. At any given time, accordingly, the working plant may include facilities ranging from earlier types to newly developed types which are barely out of the experimental stage. Before the advent of the telephone repeater, the majority of long distance facilities were open wire and, in order to keep the attenuation down, practically all of this open wire was loaded with relatively high inductance coils spaced at intervals of about 8 miles. The conductors used were almost entirely 165, 128, or 104 hard drawn copper wire and each group of four wires was usually arranged to carry a phantom circuit.

The wires were carried on crossarms in the manner indicated in Figure 18-1. Here each crossarm carries 10 wires which are numbered consecutively starting with the left-hand pin of the top crossarm when looking in the direction of the pole
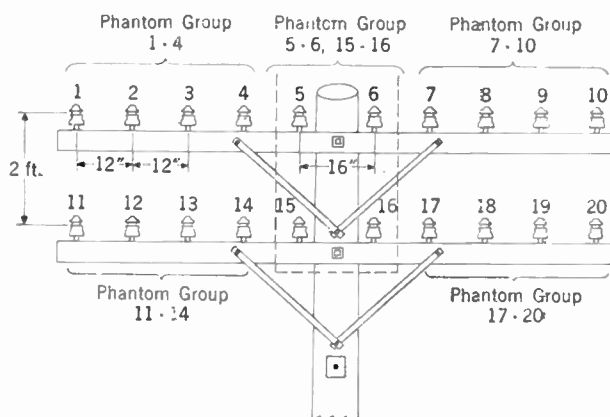
FIG. 18-1 WIRE CONFIGURATION FOR OPEN WIRE LINE CARRYING VOICE-FREQUENCY SIDE AND PHANTOM CIRCUITS

[157]

numbering of the line. The standard wire layout on two crossarms, shown in the Figure, provides ten **side** and five **phantom** circuits. Phantoms are derived from wires 1-4, 7-10, 11-14, 17-20, and 5-6, 15-16. The last is called a **vertical** or **pole-pair** phantom and has somewhat different electrical characteristics than the other phantoms because of the different spacing and configuration of the wires. Similarly the characteristics of the "non-pole-pair" side circuits such as 1-2 or 9-10, with 12 inch spacing between wires, are slightly different from those of the pole-pair circuits like 15-16, where the distance between wires is 18 inches.

Many open wire lines, with an arrangement of wires on poles as shown in Figure 18-1, are still in use in the long distance plant. Loading, however, is no longer used on open wire facilities. This is a result of the fact that the characteristics of open wire circuits—particularly the leakage—change markedly with varying weather conditions. In dry weather, open wire loading is effective in reducing the attenuation of the circuits considerably. But, due principally to the increased
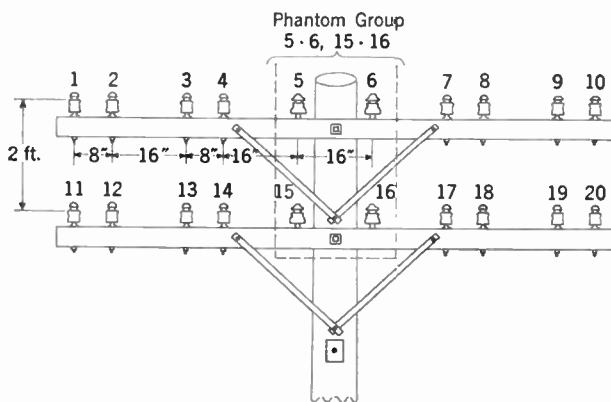
Fig. 18-2 Wire Configuration for Open Wire Line on which Type-C Carrier Systems are Superimposed

leakage, loading may actually increase the attenuation of open wire circuits in wet weather. In order to increase the overall transmission stability of such circuits, accordingly, all loading was removed after the telephone repeater came into general use, and the resulting increase in attenuation was compensated for by the employment of additional repeaters.

The application of carrier systems to open wire lines has led to other changes in open wire facility arrangements. On account of the higher frequen-
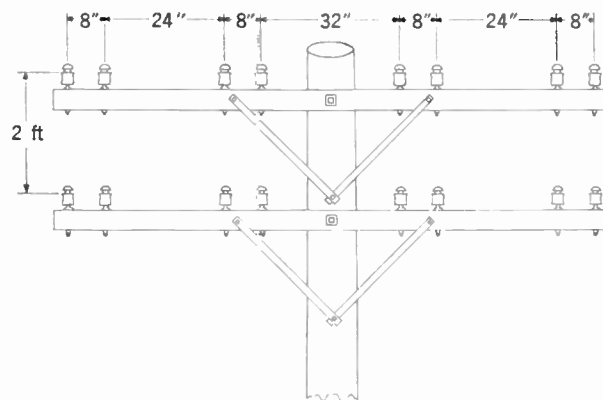
Fig. 18-3 Wire Configuration of 8-inch Spaced Open Wire Line for Type-J Carrier Operation

cies employed in carrier systems, the probability of crosstalk is increased. Since the greatest crosstalk hazard is between the side and phantom circuits of a phantom group, it is desirable in many cases to dispense with the phantom circuit altogether. Further reduction in crosstalk possibilities is effected by spacing the two wires of each pair closer together on the crossarm, and increasing the separation between pairs. Thus, Figure 18-2 shows a wire configuration used to a considerable extent on lines carrying Type-C telephone carrier systems (frequencies up to 30 kc) in which the non-pole pairs have eight inch spacing between wires and the separation between the nearest wires of adjacent pairs is 16 inches.

This configuration which is designated 8-16-8 includes a pole-pair phantom group which ordinarily would be used only for voice frequencies. The change in spacing from 12 inches to 8 inches reduces the linear inductance of the pair and in-
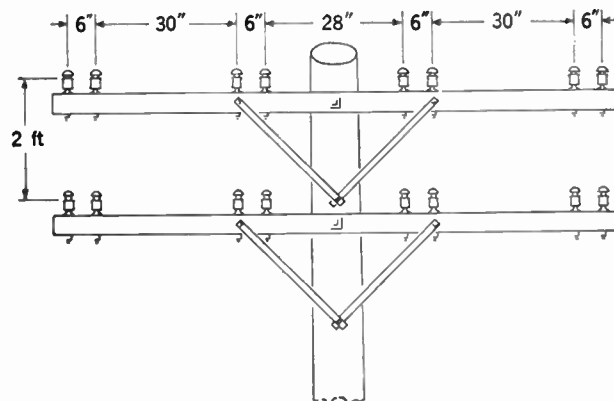
Fig. 18-4 Wire Configuration of 6-inch Spaced Open Wire Line for Type-J Carrier Operation

[158]

creases its linear capacitance by about 8%. The resistance and leakage are not changed and the attenuation is slightly increased. The characteristic impedance is reduced by about 50 ohms.

Where open wire line facilities are designed to carry broad-band carrier systems (Type-J) employing frequencies up to 140 kc, 8 or 6 inch spacing between wires of a pair is employed, and the pole-pair groups are usually dispensed with. Each crossarm then carries 8 wires, with spacings and configurations as indicated in Figures 18-3 and 18-4, and no phantom circuits are provided for. These configurations are designated 8-24-8 and 6-30-6 respectively.

Open wire facilities are subject to the effects of leakage which increase attenuation losses, particularly at carrier frequencies, and which must be adequately controlled to obtain satisfactory transmission. This is done by insulating the wires from their supporting structure with glass insulators. The effectiveness of such insulators under given conditions of weather varies with their size, shape, and the kind of pin employed.

When new open wire facilities are placed on existing lines and are likely to be used for carrier operation at frequencies above 10 kc, it is necessary to take into account the manner in which the other pairs on the line are insulated. This arises from the fact that the wet weather attenuation of similar gage facilities equipped with different types of insulators is unequal and that as a result energy level differences may occur, which cause crosstalk. When this is the case, it may be
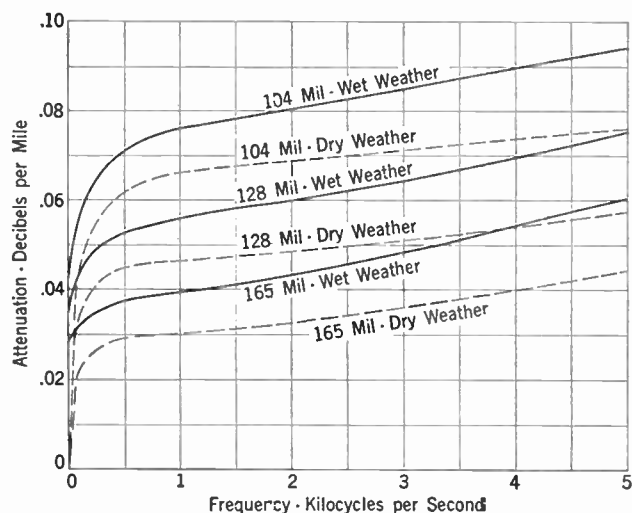
Fig. 18-6 Attenuation-Frequency Characteristics of Open Wire Phantom Circuits Over the Voice Range

desirable that all of the open wire facilities of the same gage which are to be used for carrier operation at frequencies above 10 kc be equipped with the same type insulators.

Table IX gives the more important physical and electrical constants of the commonly used types of open wire circuits. The values given are calculated for the single frequency of 1000 cycles and they apply only under more or less ideal conditions. Caution must therefore be used in applying them to practical problems. For example, the leakage of open wire conductors depends upon weather conditions. In wet weather the values for G given in the Table may be very considerably increased, and the various constants dependent to a greater or lesser extent on this value, such as attenuation, wavelength, and characteristic impedance, would change accordingly.

The Table of course does not give information regarding any variations of the circuit constants through the voice-frequency range. In practically all cases, however, the attenuation, as well as certain of the other circuit constants, changes somewhat with changing frequency. The magnitude of this attenuation change can be determined from curves in which attenuation is plotted against frequency through the working range. Figures 18-5 and 18-6 give representative attenuation-frequency curves for 104, 128, and 165 open wire, side and phantom circuits, having the wire spacing and configuration shown in Figure 18-1, over the frequency range from 0 to 5000 cycles. Separate curves are given for dry and wet weather

Fig. 18-5 Attenuation-Frequency Characteristics of Open Wire Side Circuits Over the Voice Range

TABLE IX

Characteristics of Standard Types of Open Wire Telephone Circuits at 1000 Cycles per Second

TYPE OF CIRCUIT	GAGE OF WIRES (MILS)	SPACING OF WIRES (IN.)	CONSTANTS PER LOOP MILE				PROPAGATION CONSTANT				LINE IMPEDANCE				WAVE-LENGTH MILES	VELOCITY MILES PER SECOND W	TRANS-MISSION EQUIVALENT DB PER MILE
							Polar		Rectangular		Polar		Rectangular				
			R Ohms	L Henrys	C Mf.	G Mho.	Magni-tude	Angle Degrees +	α	β	Magni-tude	Angle Degrees −	R Ohms	X Ohms −			
Non-Pole Pair Side	165	12	4 11	00337	00915	29	0252	84 36	00346	0350	612	5 35	610	57	179 5	179,500	.030
Pole Pair Side	165	18	4 11	00364	00863	29	0355	84 75	00325	0353	653	5 00	651	57	178 0	178,000	.028
Non-Pole Pair Phan.	165	12	2 06	00208	01514	58	0355	85 34	00288	0354	373	4 30	372	28	177 5	177,500	.025
Pole Pair Phan.	165	18	2 06	00207	01563	58	0359	85 33	00293	0358	366	4 33	365	28	175 5	175,500	.025
Non-Pole Pair Phys.	165	8	4 11	00311	00996	14	0353	83 99	00370	0351	565	5 88	562	58	179 0	179,000	.032
Non-Pole Pair Phys.	165	6	4 11	00292	01070	14	0356	83 63	00394	0353	529	6 25	526	58	177 8	177,800	.034
Non-Pole Pair Side	128	12	6 74	00353	00871	29	0356	81 39	00533	0352	650	8 32	643	94	178 5	178,500	.046
Pole Pair Side	128	18	6 74	00380	00825	29	0358	81 95	00502	0355	693	7 72	686	93	177 0	177,000	.044
Non-Pole Pair Phan.	128	12	3 37	00216	01454	58	0357	82 84	00445	0355	401	6 73	398	47	177 0	177,000	.039
Pole Pair Phan.	128	18	3 37	00215	01501	58	0362	82 82	00453	0359	384	6 83	382	46	174 8	174,800	.039
Non-Pole Pair Phys.	128	8	6 74	00327	00944	14	0358	80 85	00569	0353	603	8 97	596	94	178 0	178,000	.049
Non-Pole Pair Phys.	128	6	6 74	00308	01011	14	0361	80 33	00606	0356	568	9 53	560	94	176 6	176,600	.053
Non-Pole Pair Side	104	12	10 15	00366	00837	29	0363	77 93	00760	0355	692	11 75	677	141	177 0	177,000	.066
Pole Pair Side	104	18	10 15	00393	00797	29	0365	78 66	00718	0358	730	10 97	717	139	175 5	175,500	.062
Non-Pole Pair Phan.	104	12	5 08	00223	01409	58	0363	79 84	00640	0357	421	9 70	415	71	176 0	176,000	.056
Pole Pair Phan.	104	18	5 08	00222	01454	58	0368	79 81	00651	0362	403	9 83	397	69	173 6	173,600	.056
Non-Pole Pair Phys.	104	8	10 15	00340	00905	14	0367	77 22	00811	0358	644	12 63	629	141	175 5	175,500	.070
Non-Pole Pair Phys.	104	6	10 15	00322	00967	14	0371	76 60	00859	0361	610	13 26	594	140	174 2	174,200	.075

Notes: 1. All values are for dry weather conditions.
2. All capacity values assume a line carrying 40 wires.
3. Resistance values are for temperature of 20° C. (68° F.).

conditions but the latter curves naturally represent merely an average situation since the "degree of wetness" of the weather is a rather variable quantity. From these curves, it will be noted that, in general, there is an increase of attenuation between 500 and 5000 cycles of somewhere in the order of 50%.

As would be expected, when open wire circuits are used as conductors for carrier systems, the variation in attenuation from the low- to the high-frequency end of the transmission band is much greater. Thus, Figure 18-7 gives curves for 8 inch spaced, physical circuits, transposed for Type-C carrier, through the frequency range up to 50,000 cycles. Here, in the band between 5000 and 50,000 cycles, it will be seen that the attenuation more than doubles. Similarly as shown in Figure 18-8, the losses over the open wire broad-band carrier range (Type-J) increase by almost 300% in the

FIG. 18-8 ATTENUATION-FREQUENCY CHARACTERISTICS OF OPEN WIRE PHYSICAL CIRCUITS OVER THE TYPE-J CARRIER RANGE

accumulating on the wires. Thus, Figure 18-9 gives a representative example of the measured effect of melting glaze of an estimated diameter of ½ inch on an 8-inch spaced pair of 165-gage wires. Here, the attenuation at 140 kc is some four times the normal wet weather attenuation.

FIG. 18-7 ATTENUATION-FREQUENCY CHARACTERISTICS OF OPEN WIRE PHYSICAL CIRCUITS OVER THE TYPE-C CARRIER RANGE
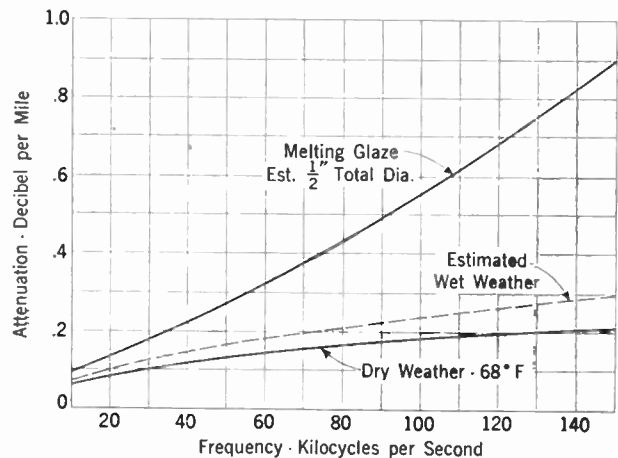
FIG. 18-9 CURVE SHOWING THE EFFECT OF SLEET DEPOSIT ON ATTENUATION OF OPEN WIRE CIRCUIT

range from 20 to 140 kc. Moreover, in the higher carrier ranges, the loss of open wire circuits may be increased to values very much larger than those indicated in this latter Figure by unusual weather conditions, such as ice, sleet or snow

[161]

18.3 Toll Cable Facilities

The use of cable conductors for long distance telephone transmission presented very considerable difficulties in the early days of the art. For obvious economic reasons, wire conductors in cables are of considerably finer gage than open wire conductors, which of course increases their attenuation per unit length. The much higher capacitance, caused by the necessary close spacing of the conductors within the cable sheath, also adds to their losses. In general, accordingly, cable conductors used for long distance voice-frequency transmission are loaded.

Before the development of the telephone repeater, toll cables were built with the largest gage conductors practicable—10, 13, and 16—and the loading was "heavy". That is to say, loading coils having inductances as high as .245 henry were inserted at intervals of 6000 to 9000 feet. As we noted in our discussion of loading in Chapter 16, however, such heavy loading, while effective in reducing the attenuation, has some undesirable effects. In the first place, it reduces the velocity of propagation to relatively low values which may seriously interfere with effective transmission over the longer circuits. Also, such a loaded circuit acts as a low-pass filter with a relatively low cutoff frequency—in the neighborhood of 2500 cycles. Requirements for high quality transmission demand that cutoff points be higher than this.

It is desirable that the ordinary cable telephone circuit transmit frequencies up to at least 3000 cycles without substantial attenuation distortion, and circuits used for program transmission work must handle frequencies much higher than this. There has been a continuing tendency, therefore, to use lighter and lighter loading in cable circuits —that is, to employ lower inductance coils and closer spacing between coils. At the same time, the general application of the telephone repeater has made it possible to use finer gage wire in toll cables, so that practically all conductors in such cables are now of either 16- or 19-gage.

The computed constants at 1000 cycles of the types of circuits extensively used in toll cables are given in Table X. This Table is similar in make-up to Table IX and the general comments made in the preceding Article regarding the former Table also apply to this. As Table X includes loading constants, however, it should be noted that the secondary constants in this case are computed on the assumption that the loading coil inductance and resistance are added directly to the corresponding basic wire constants and uniformly distributed. Certain of the resulting secondary constants, particularly the characteristic impedance, may therefore be expected to differ somewhat from values computed on the basis of "lumped" loading.

The loading designations given in the third column of Table X make use of a standardized code. The first letter in the code indicates the spacing between coils as shown in Table XI. In general, only the H and B spacings are used in toll cables proper. The other spacings listed in the Table apply either to exchange area cables or to toll entrance cables, which are discussed in the next Article. The number following the first letter in the code gives the inductance in millihenrys and the letter S, P or N following this indicates whether the circuit is side, phantom or physical (non-phantom), respectively. For example, H-25-P indicates a phantom circuit equipped with 25 millihenry coils spaced at intervals of 6000 feet. For convenience, this code system is further extended so that phantom group loading may be indicated by a letter followed by two numbers. Thus, for example, B-88-50 indicates a phantom group in which the phantom circuit is loaded with coils of 50 millihenry inductance and the side circuits are loaded with coils of 88 millihenry inductance, both spaced at 3000-foot intervals. In phantom groups, loading is usually applied to the side and phantom circuits at the same point, but this is not always true. Where the spacing is different for the side and phantom loading, two letters are used in the first symbol of the code. Thus, BH-15-15 indicates a loaded phantom group where both side and phantom coils have an inductance of 15 millihenrys, but the side circuit coils are spaced at 3000-foot intervals and the phantom circuit coils at 6000-foot intervals.

In loaded cable circuits, the amount of variation of attenuation with frequency up to frequency values fairly close to the cutoff point is relatively small. This may be seen by referring to the curves of Figure 18-10. These curves also show how the cutoff frequency takes higher and higher values as lighter loading is employed. The lightest toll cable circuit loading is used for program circuits. Figure 18-10 shows the attenuation-frequency characteristic of one such type of loading (B-22) which permits the transmission of frequencies up to about 8000 cycles without serious distortion.

TABLE X

Characteristics of Standard Types of Paper Cable Telephone Circuits at 1000 Cycles per Second

TYPE OF CIRCUIT	WIRE GAGE A.W.G.	TYPE OF LOADING	CODE NO. OF LOADING COILS	SPACING OF LOAD COILS MILES	LOAD COIL CONSTANTS PER LOAD SECTION R Ohms	L Henrys	CONSTANTS ASSUMED TO BE DISTRIBUTED PER LOOP MILE R Ohms	L Henrys	C Mf.	G μ Mho.	PROPAGATION CONSTANT Polar Magnitude	Polar Angle Degrees +	Rectangular α	Rectangular β	LINE IMPEDANCE Polar Magnitude	Polar Angle Degrees −	Rectangular R Ohms	Rectangular X Ohms	WAVE-LENGTH MILES	COILS PER WAVE-LENGTH	VELOCITY MILES PER SECOND W	VELOCITY MILES PER SECOND W	CUT-OFF FREQUENCY f_c (APPROX.)	TRANSMISSION EQUIVALENT DB PER MILE (CALCULATED)
Side	19	N.L.S.	—	—	—	—	85 8	.001	062	1 5	1830	46 98	.1249	.1338	470 1	42 80	344 9	319 37	46 93	—	—	46930	—	1 08
"	19	H-31-S	M-4	1 135	2 7	031	88 2	.028	062	1 5	2769	76 58	0643	2693	710 0	13 20	691 2	162 17	23 33	20 55	20555	23331	6700	56
"	19	H-44-S	M-2 & M-3	1 135	4 1	.043	89 4	.039	062	1 5	3188	79 87	0561	3138	818 0	9 91	805 8	140 80	20 02	17 64	17638	20022	5705	49
"	19	H-88-S	M-11	1 135	7 3	088	92 2	.078	062	1 5	4408	84 56	0418	4388	1131 0	5 22	1126 3	102 83	14 32	12 61	12615	14319	3997	36
"	19	H-172-S	M-1	1 135	13 0	170	97 3	.151	062	1 5	6095	86 96	0323	6085	1564 7	2 82	1562 8	76 90	10 33	9 10	9098	10326	2878	28
"	19	H-174-S	D-99318	1 135	16 1	.171	100 0	.152	062	1 5	6116	86 90	0331	6107	1570 0	2 84	1568 0	78 83	10 29	9 59	9586	10288	2870	29
"	19	B-88-S	M-9	0 568	7 3	.088	98 7	.156	062	1 5	6195	87 01	0322	6186	1590 2	2 76	1588 3	76 70	10 16	17 88	17882	10157	5655	28
"	16	N.L.S.	—	—	—	—	42 1	.001	062	1 5	1288	49 13	0842	0974	330 7	40 65	250 9	215 39	64 51	—	—	64506	—	73
"	16	H-31-S	M-4	1 135	2 7	.031	44 5	.028	062	1 5	2659	82 79	0334	2638	682 5	6 99	677 4	83 02	23 82	20 99	20985	23818	6700	29
"	16	H-44-S	M-2 & M-3	1 135	4 1	.043	45 7	.039	062	1 5	3148	84 61	0296	3134	808 0	5 17	804 7	72 83	20 05	17 66	17663	20048	5705	26
"	16	H-88-S	M-11	1 135	7 3	.088	48 5	.078	062	1 5	4380	87 64	0224	4374	1124 0	2 71	1122 8	53 09	14 36	12 66	12656	14365	3997	19
"	16	H-172-S	M-1	1 135	13 0	.170	53 6	.151	062	1 5	6084	88 27	0183	6082	1562 0	1 51	1561 5	11 06	10 33	9 10	9102	10331	2878	16
"	16	H-174-S	D-99318	1 135	16 1	.171	56 3	.152	062	1 5	6105	88 20	0191	6102	1567 0	1 58	1566 4	43 11	10 30	9 06	9062	10297	2870	17
"	16	B-88-S	M-9	0 568	7 3	.088	54 9	.156	062	1 5	6184	88 29	0185	6181	1587 4	1 49	1586 9	41 35	10 17	17 90	17897	10165	5655	16
Phantom	19	N.L.P.	—	—	—	—	42 9	.0007	100	2 4	1646	47 78	1106	1219	262 1	41 97	194 8	175 23	51 53	—	—	51525	—	96
"	19	H-18-P	M-4	1 135	1 4	018	44 1	.017	100	2 4	2695	78 67	0529	2642	428 8	11 11	420 8	82 61	23 78	20 95	20952	23781	6959	46
"	19	H-25-P	M-2 & M-3	1 135	2 1	025	44 7	.023	100	2 4	3082	81 30	0466	3047	490 7	8 48	485 3	72 39	20 62	18 16	18158	20621	5916	40
"	19	H-50-P	M-11	1 135	3 7	050	46 2	.045	100	2 4	4243	85 25	0351	4228	675 2	4 53	673 1	53 33	14 86	13 09	13093	14861	4193	30
"	19	H-63-P	M-1	1 135	6 1	063	48 3	.056	100	2 4	4724	85 98	0331	4712	751 8	3 80	750 1	49 81	13 33	11 75	11748	13334	3738	29
"	19	H-106-P	D-99318	1 135	8 2	107	50 1	095	100	2 4	6135	87 49	0269	6129	976 4	2 29	975 6	38 98	10 25	9 03	9033	10252	2871	23
"	19	B-50-P	M-9	0 568	3 7	050	49 4	.089	100	2 4	5939	87 37	0273	5933	945 2	2 41	944 4	39 83	10 59	18 64	18645	10590	5936	24
"	16	N.L.P.	—	—	—	—	21 0	.0007	100	2 4	1161	50 02	0746	0890	184 8	38 98	143 7	116 29	70 60	—	—	70604	—	65
"	16	H-18-P	M-4	1 135	1 4	018	22 2	.017	100	2 4	2618	84 02	0273	2604	416 7	5 76	414 6	41 83	24 13	21 26	21259	24129	6959	24
"	16	H-25-P	M-2 & M-3	1 135	2 1	025	22 8	.023	100	2 4	3032	85 41	0243	3022	482 5	4 37	481 1	36 80	20 79	18 32	18319	20792	5916	21
"	16	H-50-P	M-11	1 135	3 7	050	24 3	.045	100	2 4	4223	87 43	0189	4218	672 1	2 35	671 5	27 52	14 90	13 12	13124	14896	4193	16
"	16	H-63-P	M-1	1 135	6 1	063	26 4	.056	100	2 4	4709	87 74	0185	4705	749 4	2 04	748 9	26 63	13 35	11 77	11766	13354	3738	16
"	16	H-106-P	D-99318	1 135	8 2	.107	28 2	095	100	2 4	6128	88 54	0156	6126	975 2	1 24	975 0	21 13	10 26	9 04	9037	10257	2871	14
"	16	B-50-P	M-9	0 568	3 7	050	27 5	089	100	2 4	5931	88 48	0157	5929	943 9	1 30	943 7	21 39	10 60	18 66	18657	10597	5936	14
Physical	19	B-22	620	0 568	1 28	022	88 1	040	062	1 5	3221	80 23	0546	3174	826 8	9 55	815 3	137 20	19 79	34 85	34850	19790	11276	47
"	16	B-22	616	0 568	1 25	022	44 3	040	062	1 5	3153	84 89	0281	3140	809 4	4 89	806 4	69 01	20 01	35 26	35260	20010	11276	24

NOTE. The values for cut-off frequency and transmission equivalent per mile as given in the last two columns, are calculated from the primary constants which are assumed as uniformly distributed. These values accordingly may not be identical with the measured values given in standard formal instructions. The values given in Bell System Practices should therefore be used for engineering work.

For program transmission up to 15 kc still lighter loading is used. One system employs 11 millihenry coils spaced at 1500-foot intervals; another uses 7.5 millihenry coils spaced at 1000-foot intervals. Either system may be applied to either 16 or 19 gage conductors. Where cable circuits are used for high-frequency carrier transmission —up to a maximum of 60 kc in the case of Type-K carrier—loading is not practicable. The attenuation of the non-loaded conductors is of course very much higher, as shown by the curves of Figure 18-11, but this is offset by the use of closely spaced high-gain amplifiers.

Modern toll cables include coaxials as well as the all-wire types that we have been considering. As was noted in Chapter 16, the present standard coaxial conductor used in the Bell System has a tube diameter of .375 with a central wire .1004 inches in diameter. Early and more or less experimental types of coaxial having an outer tube diameter of .27 inches and a 13-gage central wire are also in use. Coaxial cables in present practice may be manufactured to include 2, 4, 6, or 8 tubes; and in all cases they also include a number of ordinary wire pairs or quads. In the manufacturing process some "twisting" of the tubes with respect to each other and the cable sheath occurs so that the tubes are somewhat longer than the cable. In the case of an eight tube cable this extra length is about 1.54%.
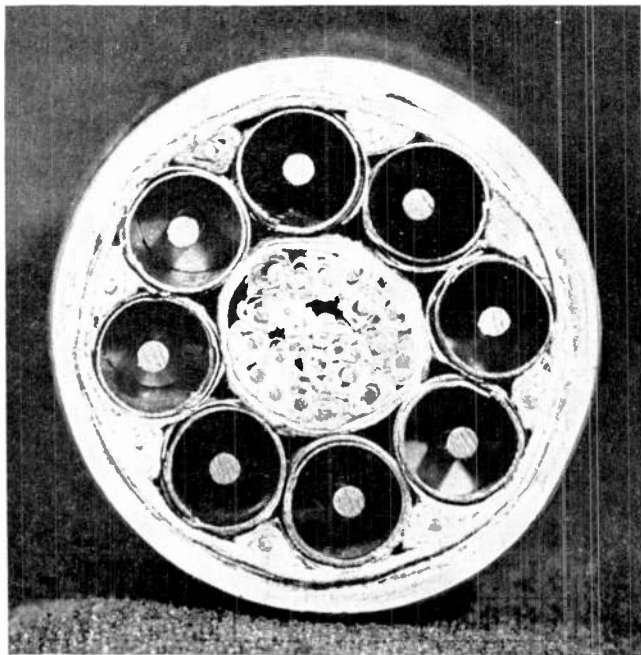
COAXIAL CABLE CROSS SECTION

TABLE XI

LOADING COIL SPACING CODE

CODE DESIGNATION	LOAD COIL SPACING IN FEET
A	700
B	3000
C	929
D	4500
E	5575
F	2787
J*	640
H	6000
X	680
Y	2130

* Spiral-four cable.

Figure 18-12 shows the attenuation loss at a temperature of 55° F of a coaxial in an 8-tube cable through frequencies ranging up to 6000 kc.

18.4 Toll Entrance and Intermediate Cable Facilities

It is seldom practicable to extend open wire line facilities into the central sections of the larger towns and cities. Instead, the open wires are terminated at a pole somewhere on the outskirts where they are connected to cable conductors extending from the terminal pole to the toll central office in the city. These terminating cables are called toll entrance cables. Depending upon the size of the city, the location of the central office, and other conditions peculiar to each particular situation, such toll entrance cables may vary in length from a few hundred feet to several miles. There are also numerous situations, such as river crossings, where it is necessary to insert relatively short lengths of cable in the long open wire line. Such cable lengths are known as intermediate cables.

In order to meet the overall transmission requirements of the long distance circuits, it is of course desirable to keep the attenuation of toll entrance and intermediate cable conductors to as low a value as practicable. It is even more important in most cases that the cable conductors should be so designed that their impedance matches the impedance of the open wire facilities to which they are connected. Loading of the proper weight is used to obtain both of these results. The cables usually contain three gages of wire—namely, 13, 16, and 19, and the larger gages are connected to the larger gages of open wire. That is, 165 open

[163]
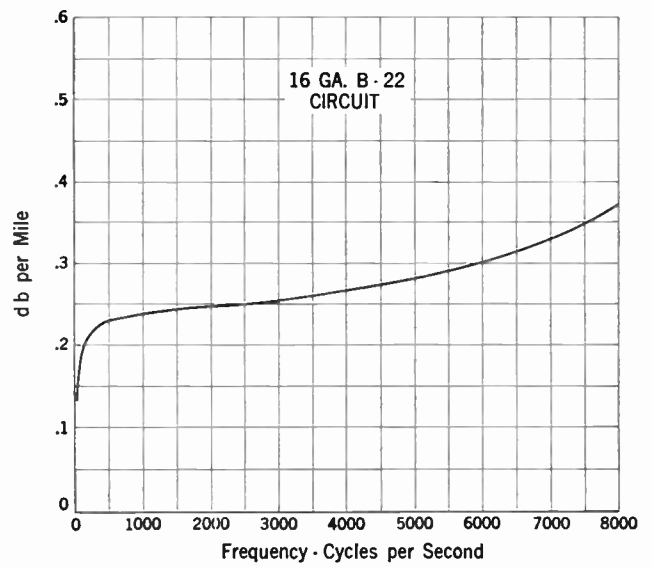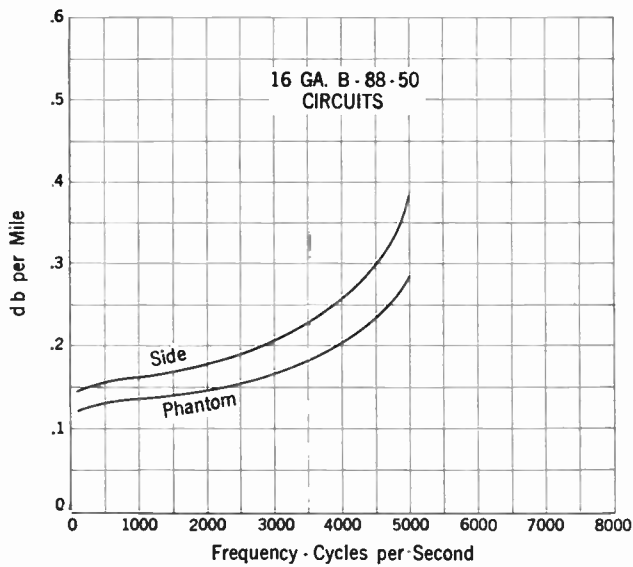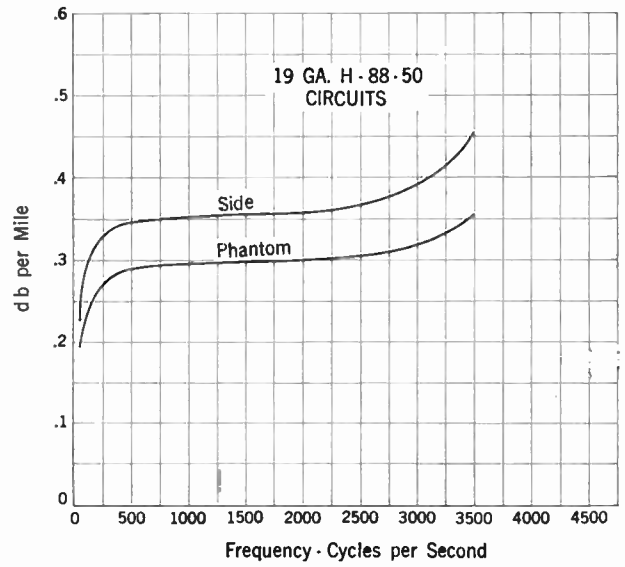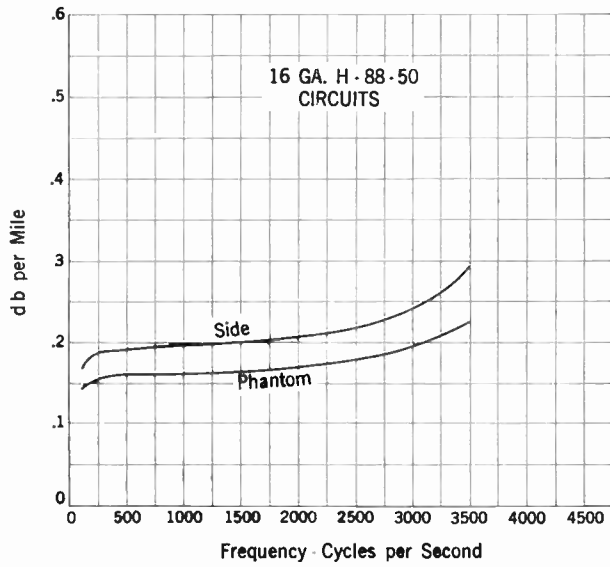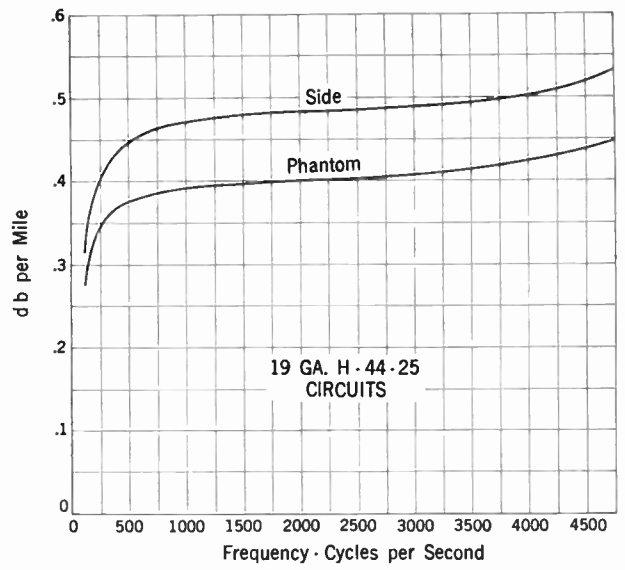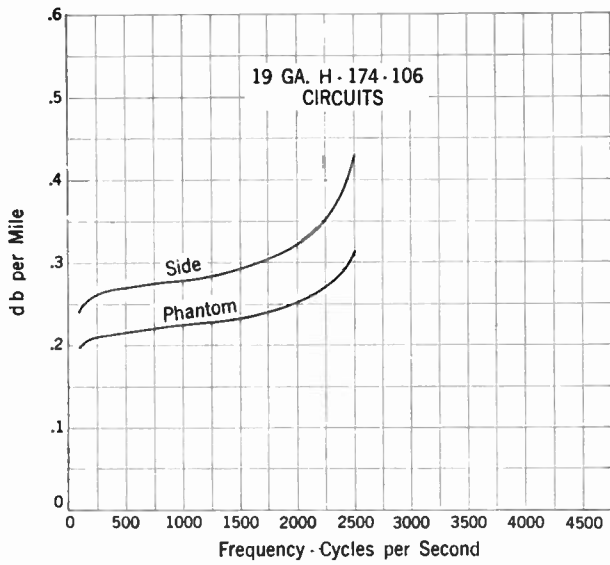
FIG. 18-10 ATTENUATION-FREQUENCY CHARACTERISTICS OF VARIOUS TYPES OF LOADED CABLE CIRCUITS

[164]
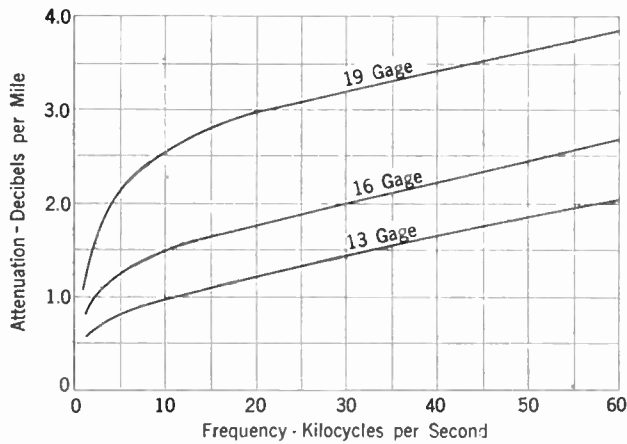
FIG. 18-11 ATTENUATION-FREQUENCY CHARACTERISTICS OF
NON-LOADED CABLE CIRCUITS OVER THE
TYPE-K CARRIER RANGE

wire will be connected to a 13-gage pair in the
cable while 104 will be connected to a 19-gage
pair. For voice-frequency open wire circuits, the
standard loading for toll entrance and intermedi-
ate cables is H-31-18. Loading may be applied to
cable lengths as short as 2000 feet, but short
lengths may require artificial building-out in order
that the loading section shall be of the proper
electrical length to give the desired characteristic
impedance.

Where open wire lines carry Type-C carrier
systems, the entrance cable loading must be very
light in order to transmit frequencies up to 30,000
cycles. For 12-inch spaced open wires, the corre-
sponding entrance cable loading generally used is
C-4.1 and C-4.8, the former being used for the
larger gage wires and the latter for the smaller.
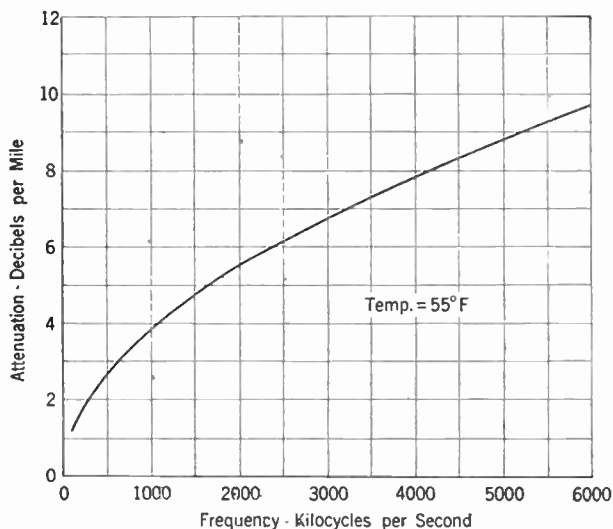
FIG. 18-12 ATTENUATION-FREQUENCY CHARACTERISTIC OF
TYPICAL .375 COAXIAL
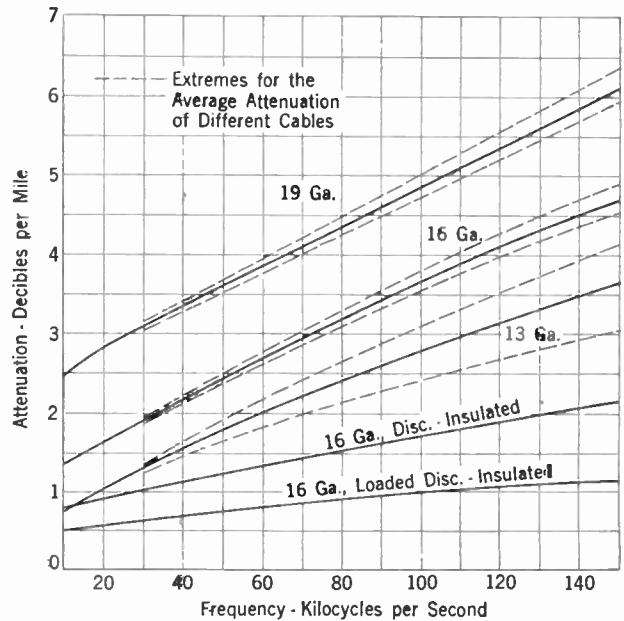
FIG. 18-13 ATTENUATION-FREQUENCY CHARACTERISTICS OF
DISC-INSULATED TOLL ENTRANCE CABLE CIRCUITS COM-
PARED WITH NON-LOADED PAPER INSULATED CIRCUITS

This loading may be modified by means of capaci-
tance and resistance building-out to be satisfac-
tory with 8-inch spaced open wire. For very short
lengths of toll entrance cable and for long lengths
of office cable used with Type-C carrier systems,
X-2.7 loading is sometimes used.

In the case of open wire lines carrying Type-J
carrier systems, the top frequencies are so high
that it is impracticable to load entrance or inter-
mediate cable conductors of the usual type prop-
erly for these systems. Non-loaded conductors are
therefore used in most cases for toll entrance and
the resulting higher attenuation is compensated
by additional repeater gain. In certain cases, par-
ticularly in intermediate cables, however, a special
type of conductor is used to handle these high fre-
quencies. This consists of a cable made up of in-
dividually shielded 16-gage disc-insulated "spiral-
four" quads. Each such quad consists of four
wires placed at the corners of a square, the two
wires at the diagonals of the square forming a
pair and having a separation of .302 inch. The
capacitance of each pair is about .025 microfarad
per mile and the attenuation (non-loaded) is about
2 db per mile at 140 kc. This may be compared
with the attenuation of ordinary non-loaded cable
pairs at comparable frequencies by referring to
Figure 18-13. These disc-insulated quads may also
be loaded to improve still further their attenua-
tion and impedance characteristics.

[165]

CHAPTER 19

ATTENUATING, IMPEDANCE MATCHING AND EQUALIZING NETWORKS

19.1 Pads or Attenuators

In the operation of various telephone circuits, it is frequently necessary to reduce artificially the currents and voltages at various points within the circuits. To accomplish this result, attenuating networks are inserted at the required points. To attenuate all currents of the different frequencies the same amount, the attenuating network obvi-

ELECTRICALLY EQUAL "T" TYPE PADS

ELECTRICALLY EQUAL "π" TYPE PADS
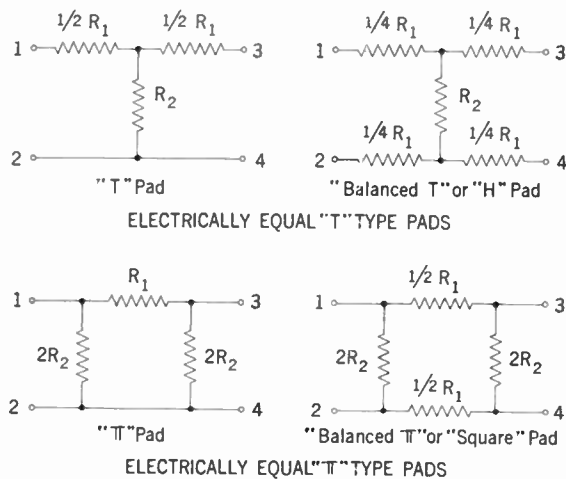
FIGURE 19-1

ously must be made up of resistances. By arranging appropriate resistances in a network of series and shunt paths, any specified value of attenuation may be obtained without introducing any impedance irregularities in the circuit in which the network is connected. Such resistance networks are usually called **pads** and the most common of these are the "T" and "π" types illustrated in Figure 19-1.

These same pads may be made up in "H" and "Square" networks where the series resistances in wires 1-3 and 2-4 are equal. This balances the two sides of the circuit without changing the electrical characteristics of the pad. Thus, in Figure 19-1, the T- and H-pads are electrically identical; that is, both pads have the same impedance (resistance) and provide exactly the same attenuation. The same applies to the π and Square pads illustrated.

In most cases pads are symmetrical; that is, their impedance, as seen from either terminals 1-2 or 3-4, is the same. This is the case for the four pads illustrated in Figure 19-1. It is possible, however, for a pad to have a different impedance as seen from either terminal. Under these conditions the pad may be used to match two unequal impedances and at the same time produce the desired attenuation.

The two basic facts required for designing a pad are, first, the impedance of the circuit in which the pad is to operate because this must match the impedance of the pad to prevent reflection loss; second, the amount of attenuation the pad is to produce. The characteristic impedance of most circuits in which pads are connected in practice is approximately a pure resistance. It is this resistance the pad is designed to match in order to prevent reflection loss.

When a symmetrical pad is terminated in its characteristic resistance, the resistance it presents at the other terminal is still, of course, its characteristic resistance. In view of this it is a simple matter to solve for the characteristic resistance of the pad in terms of its series and shunt resistances. The loss, or attenuation, of a pad is measured by the relation of the received current, I_r, to the sent current, I_s. This loss expressed in db is $20 \log_{10} \dfrac{I_s}{I_r}$. The value of this current ratio, $\dfrac{I_s}{I_r}$ can also be obtained in terms of the series,
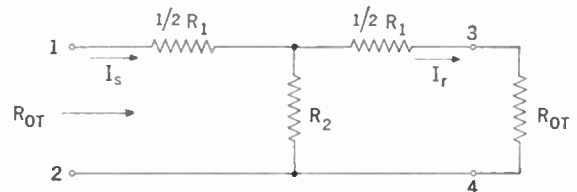
$$R_{OT} = \sqrt{\tfrac{1}{4} R_1^2 + R_1 R_2}$$

$$\frac{I_s}{I_r} = \frac{\tfrac{1}{2} R_1 + R_2 + R_{OT}}{R_2}$$

FIGURE 19-2

shunt, and characteristic resistances of the pad. These relations for both the characteristic resistance and the current ratio are given for the T- and π-pads in Figure 19-2 and 19-3 respectively.

If the same values of series (R_1) and shunt (R_2) resistances are used in T- and π-pads, their characteristic resistance and db losses will be different. However, by using proper resistances of different values, two such pads can be made electrically identical. In Article 15.4, equations (15:1), (15:2) and (15:3), show how any network can be replaced by a simple T-network, after measuring or calculating three specific resistance values. By applying these equations to the π-pad

$$R_{O\pi} = \frac{R_1 R_2}{\sqrt{1/4\ R_1^2 + R_1 R_2}}$$

$$\frac{I_s}{I_r} = 1 + \frac{R_{O\pi}}{R_2} + \frac{R_1}{2R_2} + \frac{R_1 R_{O\pi}}{4R_2^2}$$
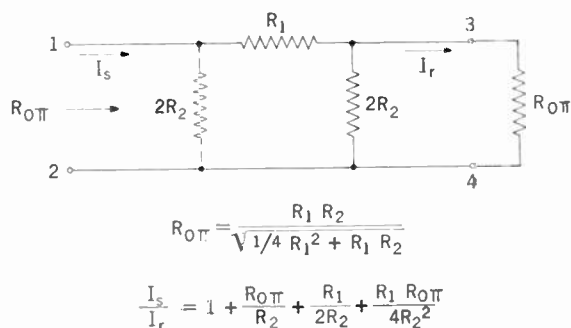
FIGURE 19-3

in Figure 19-4, the equivalent T-pad can be calculated. The resistances R_1, R_2, and R_3 in Chapter 15 are as follows:

R_1 = Resistance across terminals 1–2 with terminals 3–4 open.

R_2 = Resistance across terminals 1–2 with terminals 3–4 shorted.

R_3 = Resistance across terminals 3–4 with terminals 1–2 open.

In terms of R_x, R_y, and R_z in Figure 19-4, these resistances now become:

$$R_1 = \frac{(R_x + R_z)\ R_y}{R_x + R_y + R_z} = \frac{R_x R_y + R_y R_z}{R_x + R_y + R_z} \quad (19:1)$$

$$R_2 = \frac{R_x R_y}{R_x + R_y} \quad (19:2)$$

$$R_3 = \frac{(R_x + R_y)\ R_z}{R_x + R_y + R_z} = \frac{R_x R_z + R_y R_z}{R_x + R_y + R_z} \quad (19:3)$$

By substituting these values in equations (15:1), (15:2) and (15:3), and solving for the resistances of the T-network, we get—

$$R_a = \frac{R_x R_y}{R_x + R_y + R_z} \quad (19:4)$$

$$R_b = \frac{R_x R_z}{R_x + R_y + R_z} \quad (19:5)$$

$$R_c = \frac{R_y R_z}{R_x + R_y + R_z} \quad (19:6)$$

Where the π-pad is symmetrical ($R_y = R_z$), the values of R_a and R_b will, of course, be equal.

To convert a T-pad to a π type, we can make use of three similar equations which can be developed from equations (19:4), (19:5) and (19:6) above. This involves obtaining values of R_x, R_y, and R_z in terms of R_a, R_b, and R_c. This can be done by first obtaining the sum of the products of equations (19:4) and (19:5), (19:4) and (19:6), and (19:5) and (19:6) which gives—

$$R_a R_b + R_a R_c + R_b R_c = \frac{R_x^2 R_y R_z + R_x R_y^2 R_z + R_x R_y R_z^2}{(R_x + R_y + R_z)^2}$$

This becomes—

$$R_a R_b + R_a R_c + R_b R_c = \frac{R_x R_y R_z}{R_x + R_y + R_z}$$

Then by dividing this equation by each of equations (19:6), (19:5), and (19:4), we get—

$$R_x = \frac{R_a R_b + R_a R_c + R_b R_c}{R_c} \quad (19:7)$$

$$R_y = \frac{R_a R_b + R_a R_c + R_b R_c}{R_b} \quad (19:8)$$

$$R_z = \frac{R_a R_b + R_a R_c + R_b R_c}{R_a} \quad (19:9)$$

When the T-pad is symmetrical ($R_a = R_b$), the values of R_y and R_z will be equal.
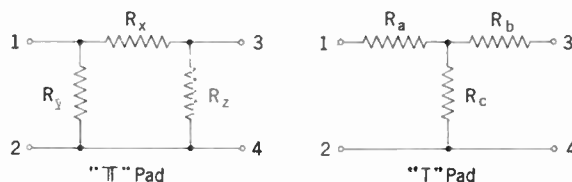
FIGURE 19-4

In designing a pad, it is only necessary to calculate the resistance values for either a T or π type and from these values, each of the types illustrated in Figure 19-1 can be obtained. Probably the simplest method is to first calculate the T-pad, and if any of the other types are desired, they can be obtained from the T.

[167]

Example: Determine the resistance values for a symmetrical 600-ohm, 10 db H-pad.

Solution: From Figure 19-2,

$$600 = \sqrt{\frac{R_1^2}{4} + R_1 R_2}$$

or

$$\frac{R_1^2}{4} + R_1 R_2 = 360,000$$

$$20 \log_{10} \frac{I_s}{I_r} = 10 \text{ db.}$$

$$\log_{10} \frac{I_s}{I_r} = \frac{10}{20} = .50$$

$$\frac{I_s}{I_r} = 3.16$$

Also from Figure 19-2

$$\frac{I_s}{I_r} = \frac{\dfrac{R_1}{2} + R_2 + R_0}{R_2} = 3.16$$

$$\frac{R_1}{2} + R_2 + 600 = 3.16 \, R_2$$

$$R_2 = \frac{\dfrac{R_1}{2} + 600}{2.16} = .231 \, R_1 + 277.8$$

Substituting R_2 in the first equation above—

$$\frac{R_1^2}{4} + R_1 \, (.231 \, R_1 + 277.8) - 360,000 = 0$$

$$.481 \, R_1^2 + 277.8 \, R_1 - 360,000 = 0$$

From this

$$R_1 = \frac{-277.8 + \sqrt{77,170 + 692,640}}{.962}$$

$$= \frac{-277.8 + 877.3}{.962} = 623.2 \text{ ohms}$$

and

$$R_2 = .231 \, R_1 + 277.8$$

$$= 144 + 277.8 = 421.8 \text{ ohms.}$$

The H-pad will have a shunt resistance (R_2) of 421.8 ohms, and each of the four series arm resistances ($R_1/4$) will be 155.8 ohms.

A common use of pads is as an auxiliary method of controlling the net gain of telephone repeaters. Such pads are usually of the square type. Table XII gives the series and shunt resistance values for symmetrical square pads of 300-ohm impedance in the range from 2.5 db to 25.0 db. Similar tables can be prepared for other

TABLE XII

SERIES (X) AND SHUNT (Y) RESISTANCE VALUES OF BALANCED π OR SQUARE PADS
$R_0 = 300$ ohms

DB Loss	Resistance Values	
	Series, X	Shunt, Y
0	0	Infinite
2.5	44	2099
5.0	91	1071
7.5	146	738
10.0	213	577
12.5	298	487
15.0	408	430
17.5	553	392
20.0	742	367
25.0	1330	336
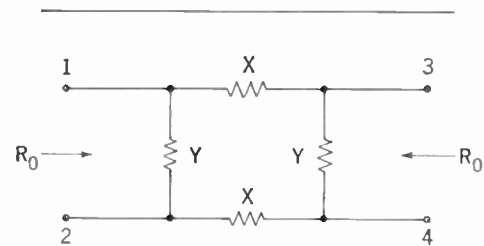
Resistance Values for Square Pads

types of pads used for various purposes in the telephone plant.

In making certain tests, it is often desirable to use a variable pad which has a fairly wide range of loss values. Such a pad is called a **variable attenuator** and usually consists of both fixed and variable H type units. The variable units are adjusted by dials while the fixed H units can be cut in or out of the attenuating circuit by a switching key. This provides a loss that can be varied over the entire range of the attenuator.

19.2 Impedance Matching

Because it can be designed to have any desired impedance values looking in either direction, a simple resistance pad can be used to match any two resistive impedances. The attenuation loss of such a pad, however, would generally be greater than the reduction in reflection loss obtained by impedance matching, although its use might still be desirable for other reasons.

Within certain limits, impedances may be matched by the insertion of a simple T- or π-pad made up entirely of inductors and capacitors. Such a pad would introduce relatively little attenuation loss but its effectiveness as an impedance matching device would of course be limited to the frequency for which its reactance values were determined. Repeating coils or transformers are therefore generally used for impedance matching purposes in voice-frequency and other relatively low frequency circuits.

As we learned in Chapter 14, however, iron-core transformers cannot be used at the higher carrier and radio frequencies because of excessive core losses. Air-core transformers also have definite shortcomings at very high frequencies. Other devices accordingly must be devised for matching unlike radio frequency transmission lines or for matching such lines to their loads (as, for example, the connection between an antenna and the transmission line from a radio transmitter). Among the most effective of such devices is the quarter-wavelength resonant line.
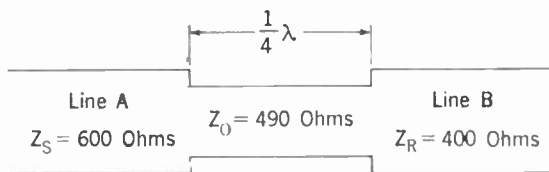
FIGURE 19-5

If an impedance Z_R is connected to one end of a quarter-wave line whose characteristic impedance is Z_0, it can be shown that at the resonant frequency (i.e., the frequency at which the line is a quarter-wavelength long) the impedance Z_S looking into the other end of the line will be—

$$Z_s = \frac{Z_0^2}{Z_R}$$

from which, we may write—

$$Z_0 = \sqrt{Z_s Z_R} \qquad (19:10)$$

A quarter-wave line designed to have the characteristic impedance Z_0 as determined from this equation will therefore serve to match two unlike impedances, Z_s and Z_R, so that there will be no reflection losses at the junction points.

As an example, assume that it is desired to connect two non-resonant lines A and B having impedances of 600 ohms and 400 ohms respectively for through transmission at a certain high fre-
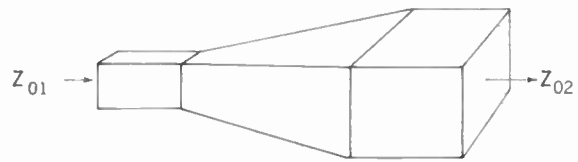
FIG. 19-6 TAPERED IMPEDANCE-MATCHING SECTION FOR WAVEGUIDE

quency. A quarter-wave line is inserted between them as indicated in Figure 19-5. Then if the quarter-wave line is designed so that its characteristic impedance $Z_0 = \sqrt{600 \times 400}$, or approximately 490 ohms, the impedance looking from line A into the quarter-wave line will be 600 ohms, and the impedance looking from line B into the quarter-wave line will be 400 ohms.

Because of this useful ability to match unequal impedances, the resonant line when so used is commonly known as a **quarter-wave transformer**. It is also interesting to note that a resonant half-wavelength line acts like a transformer of unity ratio. The impedance looking into one end of such a line is equal to the impedance connected to the other end regardless of the value of the characteristic impedance of the half-wave line itself.

Quarter-wave sections or "lines" may also be used to match waveguides of different impedances. Here the coupling section is a piece of waveguide one quarter-wavelength long with cross-sectional dimensions such that its characteristic impedance is equal to the square root of the product of the impedances of the two guides which it couples. Generally, however, matching of waveguides may be more easily accomplished by means of a simple tapered section of guide as shown in Figure 19-6. Such a section, however, must be several wavelengths long if reflection losses are to be completely avoided.

19.3 Attenuation Equalizers

One of the factors tending to decrease the intelligibility of telephone conversations is unequal attenuation of the currents of different frequencies as they pass over the circuits. For example, the attenuation of a non-loaded open wire circuit is greater for the higher frequencies than for the lower, and this difference in attenuation is directly proportional to the length of line. Therefore, when long circuits are employed it is frequently necessary to make use of **attenuation equalizers** to correct for the unequal attenuation of the line. These equalizers are usually associated with the

[169]
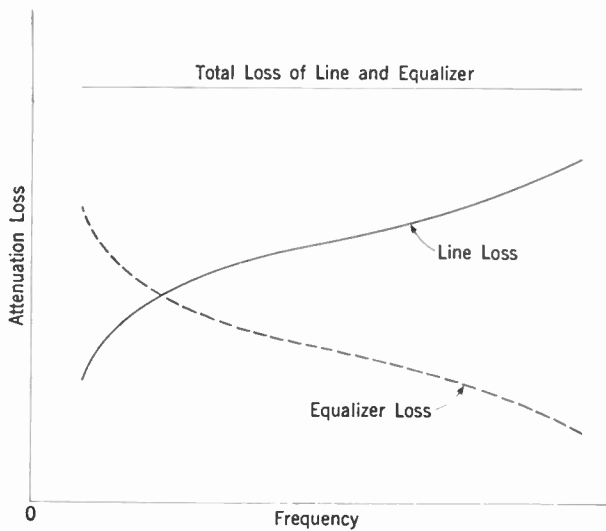
FIG. 19-7 PRINCIPLE OF ATTENUATION EQUALIZER

repeaters which must be included in the circuit to assure a satisfactory volume of sound at the receiving end.

Attenuation equalizers are networks consisting of inductors, capacitors, and resistors, which are so proportioned and arranged that their attenuation-frequency characteristics are complementary to the line characteristics that produce the distortion. In brief, the total loss of the line plus that produced by the equalizer will be substantially the same for all frequencies in the transmitted band. This principle is shown in Figure 19-7.

One of the simplest types of equalizers, shown schematically in Figure 19-8, is bridged directly across the circuit to be corrected. Obviously, the impedance of such a bridged equalizer must be low enough at certain frequencies to allow sufficient current to flow through it to produce the required losses at these frequencies. Accordingly, the equalizer circuit naturally changes the circuit impedance, particularly at the frequencies where the equalizer is to provide a substantial loss. This introduces an impedance irregularity into the circuit of a sizeable value. The use of bridged equal-
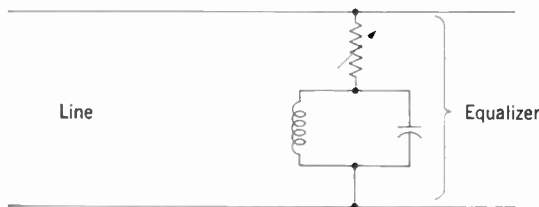
FIG. 19-8 SIMPLE BRIDGED EQUALIZER

izers of this type therefore has definite limitations in practice.

In long circuits equipped with telephone repeaters, the desired equalizing effects can be obtained without introducing an appreciable impedance irregularity by inserting equalizing networks at the mid-point of the primary sides of the repeater input transformers. Instead of changing the net loss of the line, however, this arrangement changes the overall gain-frequency characteristic of the repeater to match reasonably closely the loss-frequency characteristic of the line. That is, for the frequencies where the line loss is high the repeater gain is also high and vice versa. The overall loss-frequency characteristic of the line and repeater together is then reasonably uniform over the transmitted frequency band.

19.4 Bridged T-Equalizer

Both of the above methods of equalization give satisfactory results where the amount of attenua-
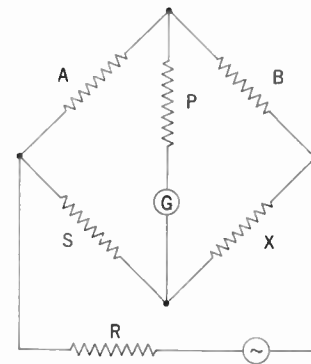
A. WHEATSTONE BRIDGE CIRCUIT
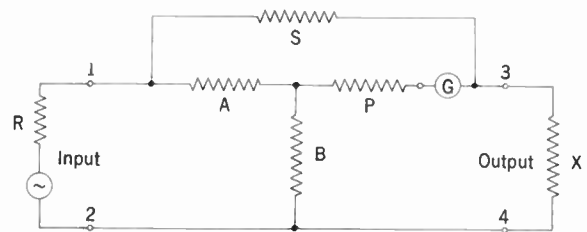
B. WHEATSTONE BRIDGE IN THE FORM OF A "T" NETWORK

FIGURE 19-9

tion distortion to be corrected is relatively small. To use either of these methods to correct a large attenuation distortion, might result in an impedance irregularity of such a magnitude as to more than offset the benefits obtained by equalizing. To equalize for these relatively large amounts of at-

[170]

CONSTANT RESISTANCE "T" NETWORK WHEN $Z_{11} Z_{21} = R^2$
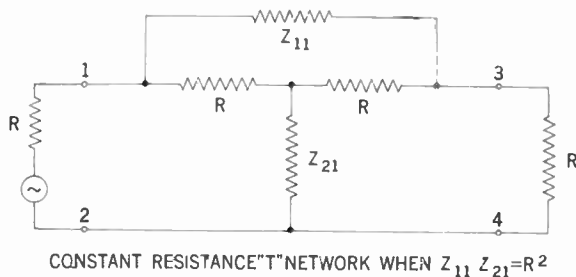
FIG. 19-10 BRIDGED T-EQUALIZER

tenuation distortion, a somewhat more complex equalizing network, in the form of a bridged T-structure, may be used. This equalizer is designed to have a constant impedance over the entire frequency band transmitted.

As its name implies, the bridged T-equalizer is built in the general form of a T-network, but it has an additional impedance path bridged across its series elements. This latter path controls the loss of the equalizer. The elements of the bridged T-equalizer are connected in a Wheatstone bridge arrangement, and the principle of its operation may be best grasped by first referring to the ordinary Wheatstone bridge circuit illustrated in Figure 19-9(A). Here a generator is connected to the two opposite points of the bridge through the impedance R, and a galvanometer, G, is connected across the other two points through an impedance P. The bridge is balanced and no current flows through the impedance P when the following proportion holds true:

$$\frac{A}{B} = \frac{S}{X} \text{ or } SB = XA$$

Now let us rearrange this bridge circuit in the form of a T-network where the series elements are bridged by the impedance S, as illustrated in Figure 19-9(B). The T-network proper is formed by A, P, and B, with S as the bridging impedance while R and X now become the input and output impedances, respectively. Next, let us change the impedances R, A, P, and X to resistances of equal value, which may then all be designated as R. For reasons to be explained later, we shall also redesignate the impedances S and B as Z_{11} and Z_{21}, respectively. Then, as illustrated in Figure 19-10, we still have the same bridge which was balanced when $SB = XA$, and is now balanced when—

$$Z_{11} Z_{21} = R \times R \text{ or } R^2 \qquad (19:11)$$

When the bridge is balanced no current flows in

the impedance, P (Figure 19-9(A)), and for purposes of analysis we may therefore simplify the network by removing the R resistance which replaced P in Figure 19-10 giving us the network of Figure 19-11. Looking from the generator across terminals 1-2 of this circuit, we now see two parallel paths which present an input impedance Z_{in} of—

$$Z_{in} = \frac{(R + Z_{11}) (R + Z_{21})}{R + Z_{11} + R + Z_{21}}$$

$$= \frac{R^2 + RZ_{11} + RZ_{21} + Z_{11}Z_{21}}{2R + Z_{11} + Z_{21}}$$

or substituting R^2 for $Z_{11}Z_{21}$,

$$Z_{in} = \frac{R(2R + Z_{11} + Z_{21})}{2R + Z_{11} + Z_{21}} = R \qquad (19:12)$$

In other words, when the bridge is balanced ($Z_{11}Z_{21} = R^2$), the input impedance of the equalizing network is a pure resistance, R. Moreover, since the T-network is symmetrical, the same reasoning can be applied at the output terminals 3-4, and the impedance will also be found to be a pure resistance, R, for the balanced condition.
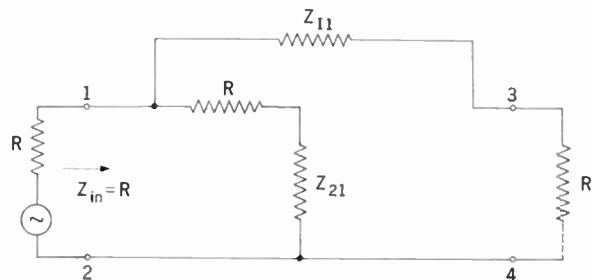
FIGURE 19-11

As in the case of any other circuit, the loss produced by this network may be determined by the ratio of the current, I_b, received in the output impedance before the network is inserted to the current, I_a, received after the network is inserted. Thus, the current, I_b, in the output before the network is inserted will be—

$$I_b = \frac{E}{R + R} \text{ or } \frac{E}{2R}$$

After inserting the network, the output of the generator will remain the same because the impedance of the network as seen at terminals 1-2 is still R. As the input current divides into the two parallel paths (Figure 19-11), the current flow-

ing in the output (terminals 3–4), I_a, is—

$$I_a = I_b \times \frac{R + Z_{21}}{R + Z_{21} + R + Z_{11}}$$

Then

$$\frac{I_a}{I_b} = \frac{R + Z_{21}}{2R + Z_{11} + Z_{21}}$$

or

$$\frac{I_b}{I_a} = \frac{2R + Z_{11} + Z_{21}}{R + Z_{21}} \qquad (19{:}13)$$

Since we are considering the balanced condition

where $Z_{11} Z_{21} = R^2$, then $Z_{21} = \dfrac{R^2}{Z_{11}}$. Substituting

this in equation (19:13), we get—

$$\frac{I_b}{I_a} = \frac{R^2 + 2R Z_{11} + Z_{11}^2}{R^2 + R Z_{11}}$$

$$= \frac{R + Z_{11}}{R} = 1 + \frac{Z_{11}}{R} \qquad (19{:}14)$$

This shows that, as long as the balanced condition is maintained, the loss of the network is determined by Z_{11}. This is also apparent from an inspection of Figure 19-11 because this impedance is in series with the receiving circuit, and any value of loss may be secured without affecting the input or output impedances, providing the balanced condition is maintained.

To summarize, **the bridged T-equalizer has a constant impedance, as seen from both terminals, equal to** R **when** Z_{21} **is the inverse of** Z_{11} **$(Z_{21} Z_{11} = R^2)$, and its overall loss-frequency characteristic is determined by the bridged series impedance network, Z_{11}.**

Both Z_{11} and Z_{21} represent generalized impedances which may be resistances, capacitances, inductances, or any combination of them. The one and only requirement is that established by the balanced condition $(Z_{11} Z_{21} = R^2)$, which means there must always be an inverse relationship between Z_{11} and Z_{21}. If Z_{11} is a pure inductive reactance represented by $j\omega L$, then Z_{21} must be $R^2/j\omega L$ or $-jR^2/\omega L$, which represents a capacitive reactance, $-j/\omega C$, where $C = L/R^2$. On the other hand, if Z_{11} is a capacitance, Z_{21} must be an inductance, which is the reverse of the above case. If Z_{11} is a resistance, then Z_{21} will also be a resistance. When Z_{11} is a network, Z_{21} is a network with the same number of elements but each element is the inverse of the corresponding element of Z_{11} as illustrated by the following table:

When Z_{11} is:	Z_{21} becomes:
Inductive reactance.	Capacitive reactance.
Capacitive reactance.	Inductive reactance.
Resistance.	Resistance.
Series inductance.	Parallel capacitance.
Series capacitance.	Parallel inductance.
Parallel resonance.	Series resonance.
Series resonance.	Parallel resonance.
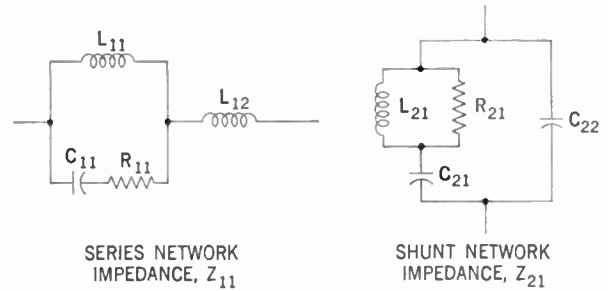
FIG. 19-12 INVERSE SERIES AND SHUNT NETWORKS

This inverse relationship is further illustrated in Figure 19-12 where the series network, Z_{11}, and its inverse shunt network, Z_{21}, are shown at left and right, respectively. Here the advantages of using the two-digit subscript for Z become more evident. The first digit of the subscript indicates whether the element belongs to the series or shunt impedance, while the second digit designates the corresponding inverse elements of the two networks. Therefore, in Figure 19-12, C_{21} is the inverse of L_{11}; C_{22} is the inverse of L_{12}; L_{21} is the inverse of C_{11}; and R_{21} is the inverse of R_{11}.

In designing a bridged T-equalizer for a specific use, the attenuation-frequency characteristic of the Z_{11} network must be complementary to the attenuation-frequency characteristic of the circuit to be corrected. This is true because, as we have seen, the loss-frequency characteristic of the bridged T-equalizer is controlled by the series impedance network, Z_{11}.

As an example of the general problem involved in the design of the Z_{11} network, let us consider a representative application of its use in a Type-C carrier system. In this, as in other carrier systems, separate frequency bands are employed for transmission in the two directions. For example, in the Type-CS system, transmission in the East to West direction occupies the frequency range from about 6 to 16 kc, while transmission in the West to East direction is in the range from about 18 to 28 kc. At the terminals and intermediate repeater points, the entire frequency band used in transmitting in each direction, which in the C systems includes three separate voice channels, is amplified by a single amplifier. The frequency

bands transmitting in opposite directions are separated by means of so-called "directional filters". The attenuation of the line facilities varies very considerably over the wide band of frequencies used. The directional filters also introduce appreciable distortion near their cutoff frequencies. In order to maintain uniform transmission, therefore, it is necessary to employ equalizers to counteract both of these attenuation distortion factors. This situation is illustrated in Figure 19-13. Here the loss produced by the line, filters, and their combined total, are indicated by the heavy lines so designated. (The frequency positions of the three voice channels in each direction of transmission are indicated by the vertical dashed lines.) The required loss-frequency characteristic of the equalizers is shown by the two
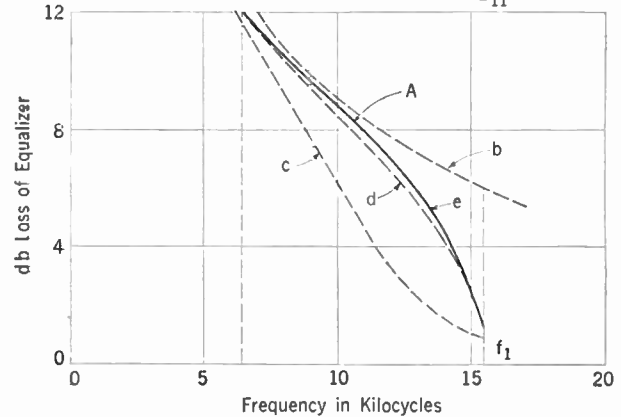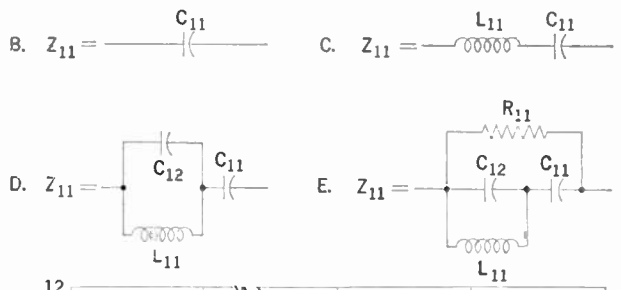
FIGURE 19-14

upper curves A and B, each of which is made complementary (inverse) to the total line and filter loss over the frequency band for its direction of transmission. By adding the losses of the line, filters, and equalizer for each direction of transmission, the resultant loss-frequency characteristic becomes a straight horizontal line in each case. Because of the rising characteristic of the line, however, the total loss for the three lower voice channels, L_1, is less than that of the three higher voice channels, L_2. This difference is readily corrected by making the amplifier gains different for the two directions of transmission.

Now that we have noted the factors which give these equalization curves (A and B) their particular characteristics, let us analyze in a general way the equalizer design considerations for one curve —say curve A. Clearly, the loss-frequency characteristic of the series impedance, Z_{11}, should conform as closely as practicable with the curve A of Figure 19-13, or with the solid line curve of Figure 19-14, which is the same. As a first approach, a Z_{11} circuit made up of a single series capacitor, as in B, will give the general loss-frequency characteristic indicated by curve b. This, of course, is due to the fact that the current through a capacitor increases with frequency; consequently, its loss decreases. However, it will be noted that curve b diverges widely from the
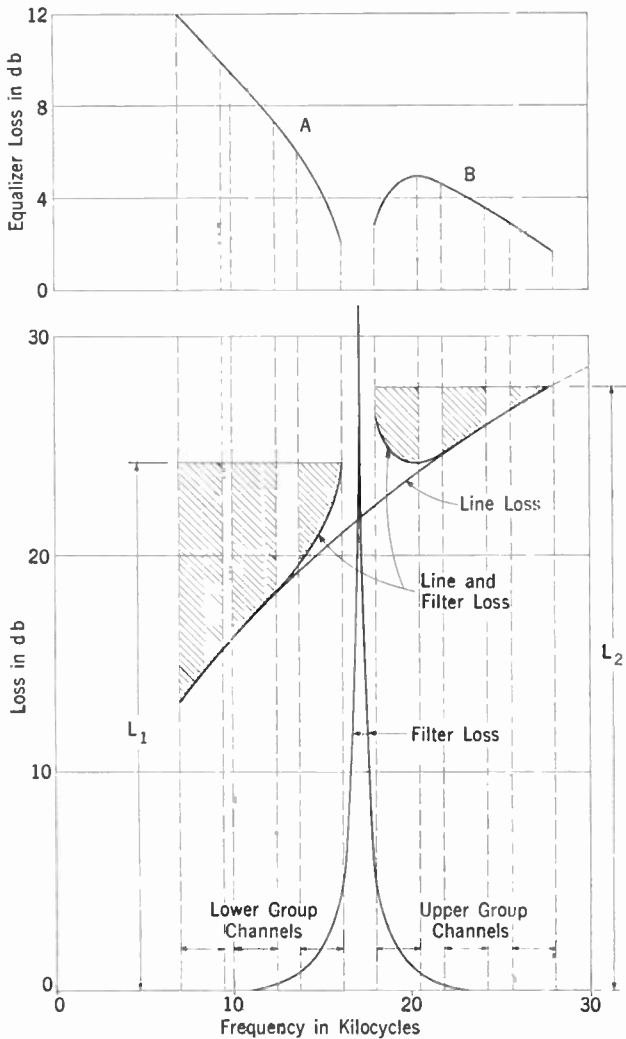
FIG. 19-13 ATTENUATION EQUALIZATION FOR TYPE-C CARRIER SYSTEM

desired characteristic at the higher frequencies. To reduce the loss at f_1, and thus bring the curves closer together, we can add in series with the capacitor an inductance, L_{11}, of such a value that series resonance occurs at approximately the frequency f_1. This fails to solve the problem, however, because below the resonant frequency this series circuit produces a loss that increases with decreasing frequency, as indicated by curve c. Because of the inductance, the curve has now become too low over most of the frequency range but yet fairly close to the desired value at the two extremities.

Apparently what is needed is an inductance that is considerably smaller than that of L_{11} over most of the frequency range but equal to it at f_1. This can be simulated by a parallel resonant circuit which has a resonant frequency somewhat above f_1, as indicated at D, because up to the resonant frequency the inductive reactance of a parallel resonant circuit increases with frequency. On
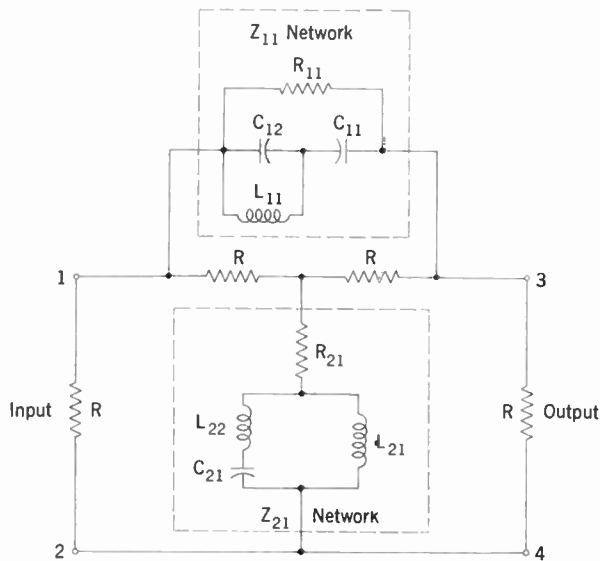
FIGURE 19-15

this basis, L_{11} can be selected so that it is small enough to approximate the desired loss at the lower and mid frequencies. Then by shunting a capacitance around it, forming a parallel resonant circuit, the effective inductance of the parallel combination at f_1 can be made equal to that of the former L_{11}. In this way the low impedance, and hence low loss, is preserved at f_1 and the loss is still increased at lower frequencies. The net effect is the characteristic shown by curve d. This comes very close to the desired characteristic, but

even greater precision can be obtained by adding the shunt resistance R_{11}, as shown by E. This introduces a small increase in the loss over most of the frequency range and modifies the curve as shown by e.
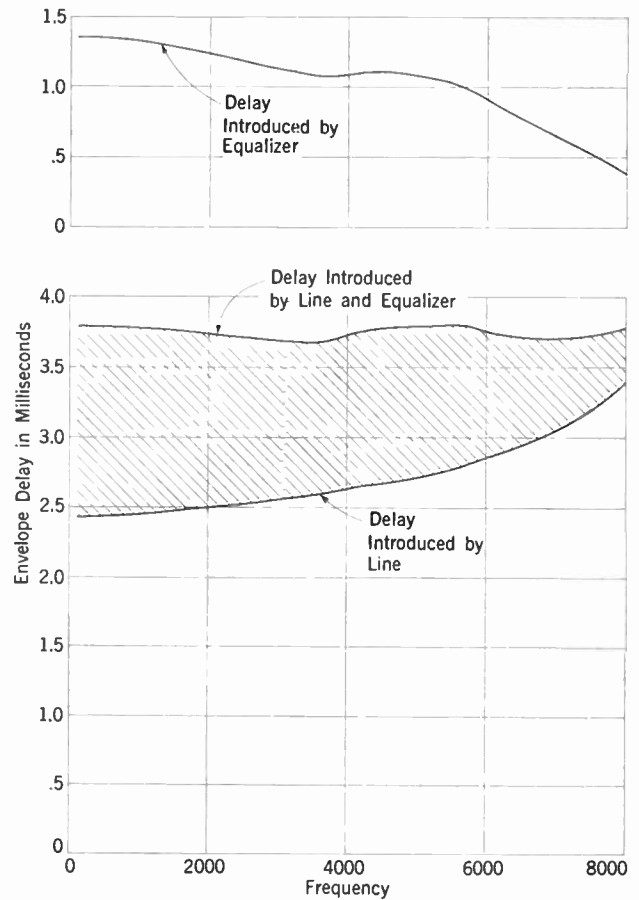
FIG. 19-16 PRINCIPLE OF DELAY OR PHASE EQUALIZING NETWORK

The final series network, Z_{11}, and its inverse, Z_{21}, then take the form illustrated in Figure 19-15. In the inverse network, Z_{21}, the shunting resistance R_{11} becomes a series resistance R_{21}; the series capacitance C_{11} becomes a shunt inductance
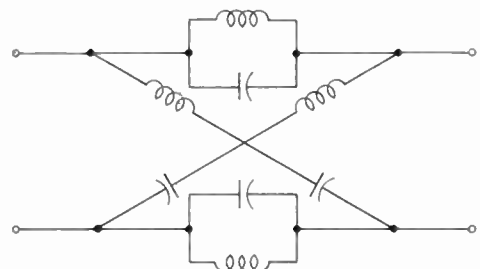
FIG. 19-17 LATTICE-TYPE NETWORK

L_{21}; while the parallel resonant circuit C_{12} and L_{11} becomes a series resonant circuit L_{12} and C_{21}. The degree of perfection with which a given loss-frequency characteristic can be matched by such an equalizer depends upon the number of coils, capacitors, or resistors it is considered economical to use.

19.5 Delay or Phase Equalizers

The equalizers discussed in the preceding Article take care of the variation of line attenuation. Unfortunately, this is not the only way in which the transmission characteristics of long telephone lines vary with frequency. The velocity of propagation, W, over these lines also may change with frequency.

A pair of wires of zero resistance in free space, separated from all other conductors and without leakage, would transmit electric waves at the speed of light. In an open wire circuit, what retardation exists comes largely from the inductance of the wires. In loaded cable circuits, there is a much greater retardation due in part to the larger capacitance between the wires, but much more to the inductance of the loading coils which are inserted to decrease the attenuation. In any case, there is a finite time interval for transmission between the sending end of any circuit and its receiving end.

This delay interval is greater with some types of facilities than others and, of course, increases in direct proportion with the length of circuit in every case. Moreover, the delay may not be the same for all frequencies in the transmitted band. A changing velocity of propagation at the various frequencies means that the shape of a signal wave at the receiving end of the line will differ to an appreciable degree from the waveshape applied at the sending end. Distortion of this kind is called **envelope distortion**, and its seriousness is measured in terms of **envelope delay** in millisec-
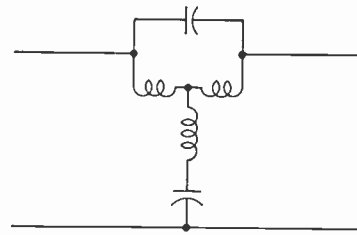
FIG. 19-18 BRIDGED T-NETWORK

onds. This delay distortion may be equalized by inserting in series with the circuit a network having the inverse characteristics. The total delay produced by the circuit, added to that of the equalizer, will then be constant over the frequency band transmitted. This is illustrated by Figure 19-16.

Delay distortion, unless excessive, is of little concern in ordinary telephone circuits because the ear seems insensitive to it. For program transmission, however, where frequency bands extending up to 8,000 or 15,000 cycles may be employed, correction for phase distortion is generally necessary. It is even more necessary in the case of facsimile (telephotograph) transmission and video transmission, where even slight delay deviations may affect the quality of the received pictures.

The design of delay equalizers involves some rather complex network analysis. Two types of networks are generally used. The first known as a balanced lattice-type structure is shown schematically in Figure 19-17. The second is a bridged-T type of structure as illustrated in Figure 19-18. In either case the total equalizer is made up of a number of such sections connected in tandem. The bridged-T type of network is somewhat more economical when balance between the two sides of the line is not necessary, although either type will produce generally similar envelope delay characteristics.

CHAPTER 20

FILTERS

20.1 Filter Requirements

In telephone and telegraph work, it is often desirable to suppress or eliminate currents of certain undesired frequencies and at the same time to pass currents of other frequencies. This is accomplished by means of **electric wave filters**. As would be expected from our study of resonance in Article 13.4, these filters are essentially networks of inductances and capacitances. While the details of design of some of the more elaborate types of filters are beyond the scope of this text, we may consider here the general principles of the four major types commonly used. These are known respectively as (1) "low-pass" filters which transmit, with very little attenuation, currents of all frequencies from zero up to some designated cutoff frequency and offer very high attenuation to all higher frequencies; (2) "high-pass" filters which perform the reverse of this action and attenuate up to the cutoff value but readily transmit currents of all higher frequencies; (3) "band-pass" filters which have both an upper and a lower cutoff point, and which permit the transmission of only those frequencies lying between the two cutoff frequencies; (4) "band-elimination" filters which have over-lapping upper and lower cutoff points and thus prevent the transmission of frequencies between the two cutoff frequencies.

At the beginning it is convenient to consider the requirements of an ideal filter, even though the ideal cannot be fully obtained in practice. In such a filter, (1) frequencies lying within the pass bands would be transmitted without hindrance; in other words, over these bands the attenuation would be zero and no power would be dissipated by the filter; (2) the frequencies outside of the pass bands would be completely suppressed and the attenuation would be infinite; (3) the frequency intervals between the transmitted and attenuated frequency bands would be very small; in other words, the change from passing to suppressing or vice versa, would occur in a very narrow transition band; (4) throughout the transmitted bands, the characteristic impedance at the

filter terminals would match the impedance of the terminating apparatus to prevent reflection losses.

From our study of resonance and the effect of resistance on the attenuation-frequency curves of resonant circuits, it is apparent that an ideal filter must be constructed entirely of pure reactances because the presence of resistance would produce attenuation in the transmitted bands. If there are no limitations as to the complexity of the reactance arms used in forming filters, or in the configuration in which these arms may be arranged, then there are an infinite number of possible designs of filters. Naturally, however, filters are designed to meet specific requirements using the simplest networks practicable.

20.2 Low- and High-Pass Filter Sections

The simplest arrangements of elementary filter networks are the T and π sections shown in Figure 20-1, where Z_1 and Z_2 represent the series and shunt impedances, respectively. (Note: the values of Z_1 and Z_2 are not necessarily the same in the two drawings.) For simplicity our discussion will be confined to the T type of network, but as we learned in the preceding Chapter, these two networks may be interchanged if certain definite
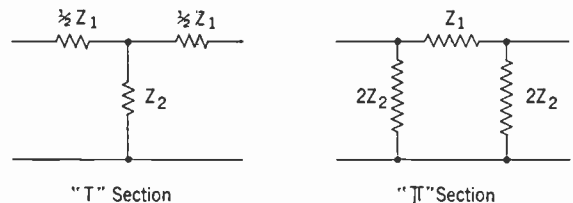
"T" Section   "π" Section

FIGURE 20-1

electrical relations are maintained. These simple networks are called **prototype** filter sections, and are the basic structures from which practical filters are developed.

As covered in Article 15.5, when a network is terminated in its characteristic impedance, Z_0, the impedance presented at the input terminals is still Z_0. Its value may be determined by taking the

square root of the product of the impedances (geometric mean) from one end with the other end open-circuited, Z_{oc}, and then short-circuited, Z_{sc}, ($Z_0 = \sqrt{Z_{oc}Z_{sc}}$). Since we are now considering an ideal filter made up of pure reactances in the form of a T-network, these open- and short-circuited impedances must be either positive (inductive), or negative (capacitive) reactance. Accordingly, the characteristic impedance must have an angle of $0°$ or $\pm 90°$. When the open- and short-circuited impedances (reactances) have opposite signs—

$$Z_0 = \sqrt{(\pm jX_{oc})(\mp jX_{sc})}$$
$$= \sqrt{-j^2(X_{oc}X_{sc})}$$

or

$$Z_0 = \sqrt{X_{oc}X_{sc}} \tag{20:1}$$

This is an impedance with a zero angle, or effectively a pure resistance. If, however, the signs are alike—

$$Z_0 = \sqrt{(\pm jX_{oc})(\pm jX_{sc})}$$
$$= \sqrt{j^2(X_{oc}X_{sc})}$$

or

$$Z_0 = \pm j\sqrt{X_{oc}X_{sc}} \tag{20:2}$$

and the impedance has an angle of $\pm 90°$ which, of course, represents a pure reactance.

When, as in the first case, the characteristic impedance is a resistance equal to the impedance of the terminating device, the network (consisting of filter and termination) is capable of absorbing power (energy) from any source connected to the input. Since none of this power can be dissipated in the reactances of the filter, it must be passed on to the receiving device (termination). In other words, the termination absorbs all the power and the attenuation in the filter itself is zero. On the other hand, when the characteristic impedance is a pure reactance, no power can be absorbed by the network. Under this condition the filter network would merely take energy from the source during part of a cycle, store it in the electric and magnetic fields, and return it during a later part of the cycle. In other words, the filter will not transmit any energy and therefore may be said to have infinite attenuation.

Over a range of frequencies, the open- and short-circuited impedances of the filter change from positive to negative and vice versa, as may be noted from Figure 20-2(B) for the three proto-type filter sections illustrated. At the same time, the characteristic impedance changes from resistance to reactance and vice versa. **When the characteristic impedance is a resistance, the filter transmits; when the characteristic impedance is a reactance, the filter attenuates.** This is also shown by the solid line curve of Figure 20-2(D).

As we have already seen, the characteristic impedance, Z_0 is

$$Z_0 = \sqrt{Z_{oc}Z_{sc}} \tag{20:3}$$

but we also know from equation (15:4) that—

$$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1 Z_2} \tag{20:4}$$

which may also be written as—

$$Z_0 = \sqrt{Z_1\left(\frac{Z_1}{4} + Z_2\right)} \tag{20:5}$$

When the reactance represented by Z_1 is opposite in sign to the reactance represented by the quantity $\left(\frac{Z_1}{4} + Z_2\right)$ their product is positive and the characteristic impedance is a resistance. On the other hand, if these two reactances are of the same sign the characteristic impedance is a pure reactance. In the first case, the filter transmits; in the second case, the filter attenuates.

The passed and attenuated frequencies may also be determined from reactance curves for Z_1 and $\left(\frac{Z_1}{4} + Z_2\right)$ as shown in Figure 20-2(C). When these two curves have opposite signs, the characteristic impedance of the filter is a resistance and the filter transmits, but when the signs are alike, the characteristic impedance is a reactance and the filter attenuates. This is illustrated by the solid curve of Figure 20-2(D). At the critical point where the $\left(\frac{Z_1}{4} + Z_2\right)$ curve crosses the zero axis, the characteristic impedance, Z_0, becomes zero as is evident from equation (20:5). This is the frequency at which the filter is said to cut off. On one side of this point is the pass band and on the other side the filter attenuates. This cutoff frequency, f_c, may be determined from the equation

$$\frac{Z_1}{4} + Z_2 = 0 \tag{20:6}$$

[177]
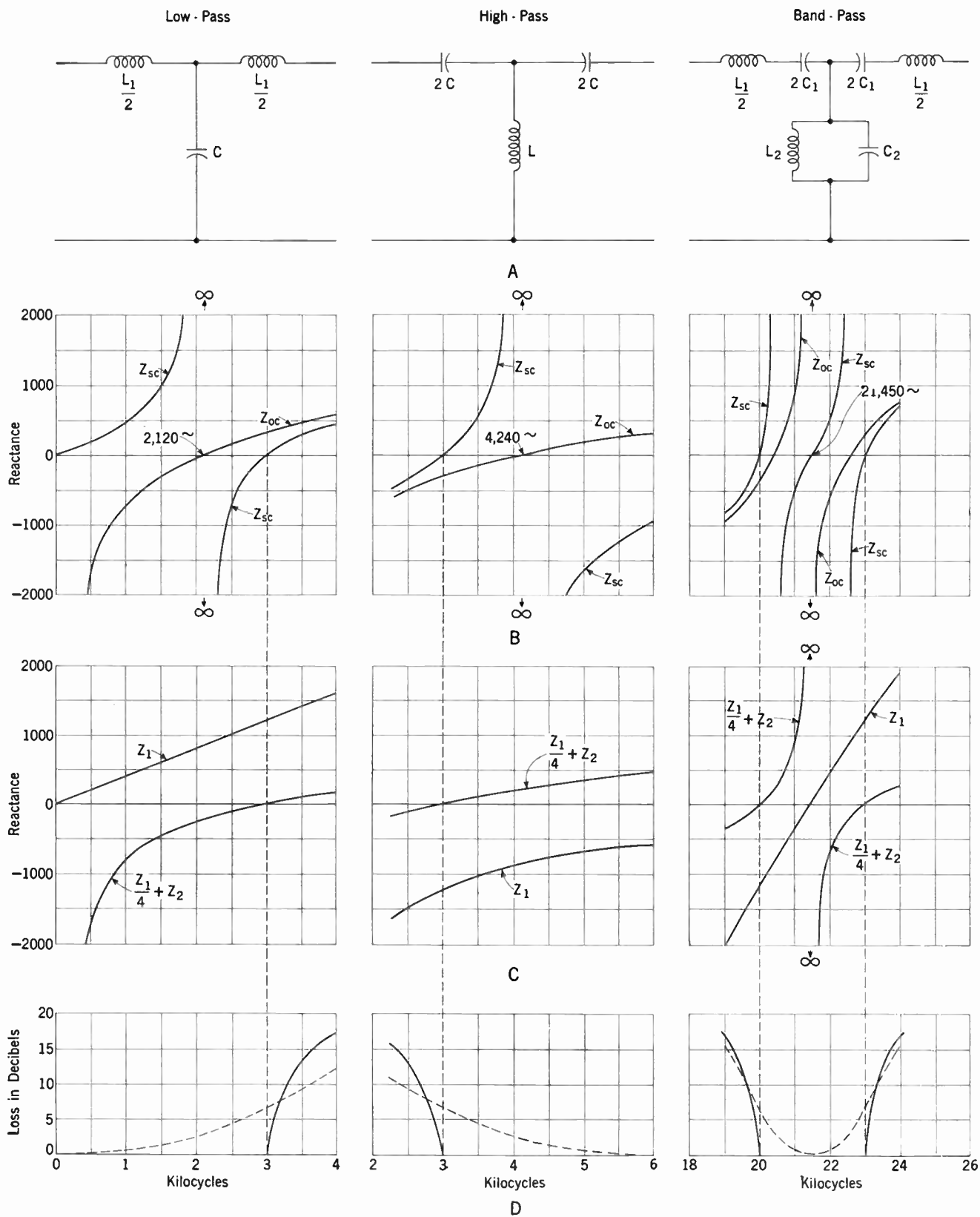
FIG. 20-2 Reactance and Loss Characteristics of Elementary Prototype Filter Sections

[178]
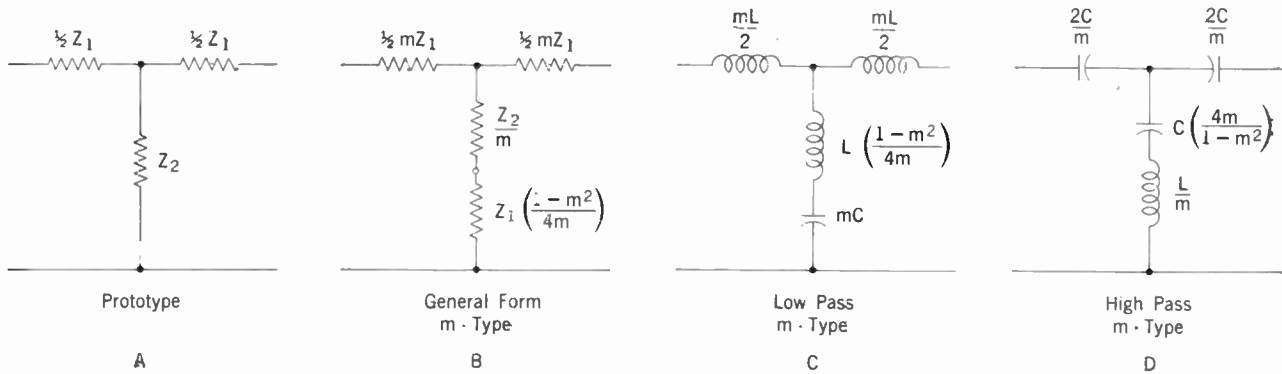
FIG. 20-3 M-TYPE FILTER SECTIONS

In the low-pass prototype filter section of Figure 20-2(A)—

$$Z_1 = j2\pi f_c L \text{ and } Z_2 = \frac{-j}{2\pi f_c C}$$

at the critical frequency, f_c. Substituting these values in equation (20:6), we get—

$$f_c = \frac{1}{\pi\sqrt{LC}} \qquad (20:7)$$

Similarly, in the high-pass prototype filter section of Figure 20-2(A)—

$$Z_1 = \frac{-j}{2\pi f_c C}, \text{ and } Z_2 = j2\pi f_c L$$

Then the cutoff frequency is—

$$f_c = \frac{1}{4\pi\sqrt{LC}} \qquad (20:8)$$

20.3 M-Derived Filter Sections

Now at this point it will be noted that although we have used the elementary, or prototype, filter structures of Figure 20-2(A) to derive equations for the cutoff frequency and to determine the pass and stop bands, as a practical matter these simple structures do not even approach the ideal filter requirements set up in Article 20.1. This will be apparent from a glance at the loss-frequency curves shown by the dashed lines in Figure 20-2(D), which illustrate the actual losses produced by each structure when inserted in a circuit of constant impedance (resistance). Because the impedances of the prototype sections do not remain constant over the frequency band, the "insertion" losses they produce depart rather widely from the theoretical losses of the ideal structures (as indicated by the solid lines) and

obviously do not meet the ideal requirements. The attenuation in the stop band is much too low, and the cutoff frequency is anything but sharp.

Some improvement could be obtained by connecting several of these structures in tandem, but the cutoff point would still be not clearly defined and there would be altogether too much attenuation in the pass band. To improve the sharpness of cutoff, what we need is a structure which will produce very much higher losses at frequencies just beyond the cutoff frequency. This objective can be approached by making the shunt impedance of the structure resonant at a frequency a few cycles beyond the cutoff frequency. Such a structure would, of course, have to contain series capacitance and inductance in its shunt arm. More complex structures containing the desired shunt resonant paths may be derived from the simple prototype structures which we have considered up to now. These derived filter sections are called *m*-**derived types** or just *m*-types, where the term *m* is a constant factor used to maintain certain relationships between the prototype and its derived type.

Let us consider a T-section as shown in Figure 20-3(A) where—

$$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1 Z_2} \qquad (20:4)$$

We wish to obtain a similar derived structure having the desired resonance characteristic in the shunt arm and whose characteristic impedance, Z'_0, is equal to the Z_0 of the prototype. (The pass and stop bands of the derived structure must, of course, be the same as those of the prototype.) Such a structure can be obtained if the impedance of the series arm Z'_1 is made equal to the series arm impedance, Z_1, of the prototype, multiplied

[179]

by a constant factor m. In other words—

$$Z'_0 = Z_0 \qquad (20:9)$$

and

$$Z'_1 = mZ_1 \qquad (20:10)$$

Then using equations (20:4), (20:9) and (20:10), we can solve for the shunt arm impedance, Z'_2, of the derived structure in terms of Z_1, Z_2, and m to get—

$$Z'_2 = \frac{Z_2}{m} + Z_1 \frac{(1 - m^2)}{4m} \qquad (20:11)$$

The m-type section then takes the general form shown in Figure 20-3(B). This applies to both low- and high-pass types but Figures 20-3(C) and 20-3(D) show the respective values in terms of m, L and C for low-pass and high-pass T-sections. It may be noted that the second term of equation (20:11) can be physically realized only if the value of m lies between zero and ± 1. By using different values of m within these limits, any number of m-type sections having the same characteristic impedance may be derived. It is only necessary to determine the particular value of m that must be used in order to provide an infinite loss a few cycles above or below the desired cutoff frequency, depending upon whether it is a low- or high-pass filter. The loss in either case will be infinity at the resonant frequency of the shunt arm, f_r; in other words, when the reactances in the shunt arm are equal in magnitude but opposite in sign. In this case—

$$\frac{Z_2}{m} + Z_1 \left(\frac{1 - m^2}{4m}\right) = 0 \qquad (20:12)$$

In the low-pass filter, we have already noted that—

$$Z_1 = j2\pi f_r L \text{ and } Z_2 = \frac{-j}{2\pi f_r C}$$

By substituting these values in equation (20:12) and solving for the frequency of infinite attenuation, f_r, we get—

$$f_r = \frac{1}{\pi\sqrt{LC(1 - m^2)}} = \frac{1}{\pi\sqrt{LC}} \times \frac{1}{\sqrt{1 - m^2}} \qquad (20:13)$$

Since, as shown in equation (20:7), the cutoff frequency in the low-pass filter is—

$$f_c = \frac{1}{\pi\sqrt{LC}}$$

m_{lp} (for the low-pass filter) can be determined in terms of f_c and f_r as—

$$m_{lp} = \sqrt{1 - \left(\frac{f_c}{f_r}\right)^2} \qquad (20:14)$$

In a similar manner, m_{hp} (for the high-pass filter) can be determined by substituting the proper values for Z_1 and Z_2 in equation (20:12), and its value found to be—

$$m_{hp} = \sqrt{1 - \left(\frac{f_r}{f_c}\right)^2} \qquad (20:15)$$

The closer the values of the two frequencies, f_c and f_r, are to each other the lower the value of m.

Curve B of Figure 20-4 is an attenuation-frequency characteristic of a representative m-derived low-pass filter section designed for a resonant frequency of 3100 cycles ($m = .252$). This may be compared with the curve for the corresponding prototype section shown as A in the same figure.

20.4 Composite Filters

A complete practical filter, frequently called a "composite" filter, consists of a prototype section connected in tandem with sufficient m-type sections to produce the desired narrow transition bands and the required loss in the stop bands. All of these sections have the same characteristic impedance since this was one of the basic factors on
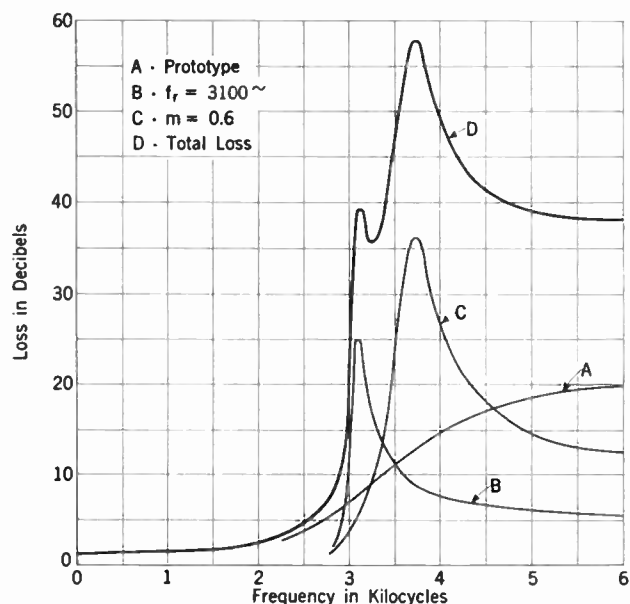
Fig. 20-4 Characteristics of Sections of Composite Low-pass Filter Shown in Fig. 20-6

[180]
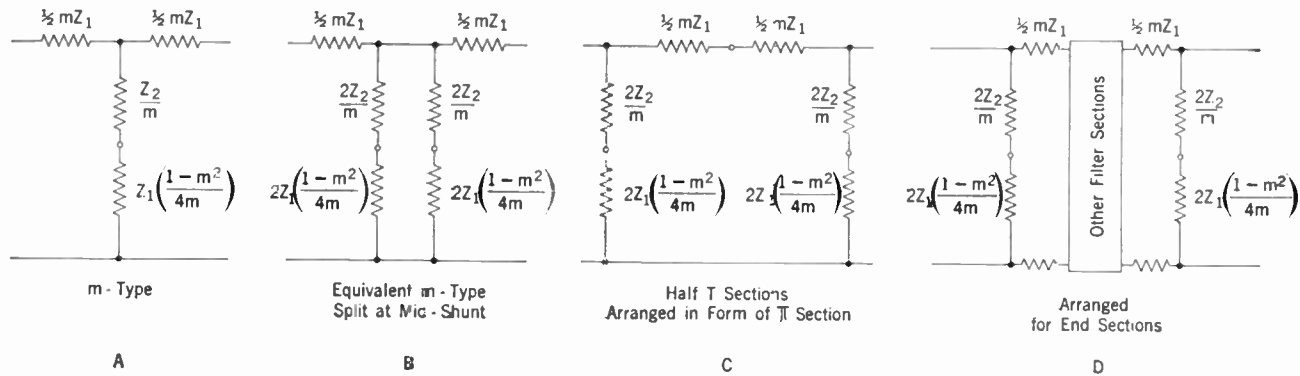
FIG. 20-5 HALF T M-TYPE FILTER END SECTIONS

which the m-type was derived from the prototype. Unfortunately, however, the characteristic impedance of T-prototypes and their m-type sections is not constant for all frequencies in the pass band. Instead, this impedance decreases rapidly and becomes zero at the cutoff value as we have seen (Figure 20-2). Such an impedance characteristic is, of course, unsatisfactory since it would cause large reflection losses when the filter was connected into a circuit of constant impedance. It therefore is highly desirable to improve the terminal impedance of the complete filter over the pass band. This can be effected by adding a "half T-section" at each end of the filter structure.

By splitting a T-section through its shunt arm, we obtain two "half T-sections", each having a shunt impedance of twice that of the original T-section, while the series arm of each of the half sections contains one half of the original total series impedance. This splitting arrangement is illustrated in Figure 20-5(B). We may then rearrange the series and shunt arms so that the network takes the form of a π as in Figure 20-5(C). It should be noted, however, that we are not converting the T-section into an electrically equivalent π-section, but simply rearranging the positions of the series and shunt arms without changing their respective values. The new structure will have the same loss as it had in its original T form but its impedance will be changed because we are now viewing the network "mid-shunt" instead of "mid-series".

Now, if characteristic impedance is plotted against frequency for various values of m it will be found that the impedance of such a structure is practically uniform over the major portion of the pass band when m is equal to 0.6. We can take advantage of this fact by breaking this π-section

into two equal parts, each designed with m equal to 0.6. and using each half as an end section of the composite filter as shown in Figure 20-5(D). The terminal impedances of the complete filter will then be practically constant over the major portion of the pass band.

In designing a filter we must know the frequency band to be passed, the cutoff frequencies, and the impedance of the circuit in which the filter is to work. In practically all cases this impedance will have a zero angle, and it may therefore be considered as a pure resistance, R. It can be shown that in such a case the matching impedance of the filter is—

$$R = \sqrt{\frac{L}{C}} \qquad (20:16)$$

for both the low- and high-pass filters. Combining this relation with the value of the cutoff frequencies as given in equations (20:7) and (20:8), we have for the low-pass filter—

$$L = \frac{R}{\pi f_c} \qquad (20:17)$$

and

$$C = \frac{1}{\pi f_c R} \qquad (20:18)$$

and for the high-pass filter—

$$L = \frac{R}{4\pi f_c} \qquad (20:19)$$

and

$$C = \frac{1}{4\pi f_c R} \qquad (20:20)$$

These are the values of the inductance and capacitance in the prototype sections. Constants of the other sections are developed from these proto-

[181]

type constants.

Example: Design a low-pass filter with a cutoff frequency of 3000 cycles to work into an impedance of $600/0°$. To provide a sharp cutoff, f_r should be 3100 cycles.

Solutions: From equations (20:17) and (20:18), the inductance and capacitance for the prototype will be—

$$L = \frac{R}{\pi f_c} = \frac{600}{\pi 3000} = .06366 \text{ henry or } 63.66 \text{ mil hen.}$$

$$C = \frac{1}{\pi f_c R} = \frac{1}{\pi 3000 \times 600}$$
$$= .1768 \times 10^{-6} \text{ farad or } .1768 \,\mu\text{f.}$$

From equation (20:14) and Figure 20-3(C), the m-type section for $f_r = 3100$ will be—

$$m_{lp} = \sqrt{1 - \left(\frac{f_c}{f_r}\right)^2} = \sqrt{1 - \left(\frac{3000}{3100}\right)^2} = 0.252$$

$$\frac{mL}{2} = \frac{.252 \times 63.66}{2} = 8.02 \text{ mil hen.}$$

$$L\left(\frac{1-m^2}{4m}\right) = 63.66\left(\frac{1-(.252)^2}{4 \times .252}\right)$$
$$= 59.16 \text{ mil hen.}$$

$$mC = .252 \times .1768 = .04455 \,\mu\text{f.}$$

In a similar manner the end sections, where $m = 0.6$, will be—

$$\frac{mL}{2} = \frac{.6 \times 63.66}{2} = 19.10 \text{ mil hen.}$$

$$L\left(\frac{1-m^2}{4m}\right) = 63.66\left(\frac{1-(.6)^2}{4 \times .6}\right) = 16.98 \text{ mil hen.}$$

$$mC = .6 \times .1768 = .106 \,\mu\text{f.}$$

The composite filter thus takes the form illustrated in Figure 20-6. The attenuation-frequency characteristics for this composite filter and each of its component structures are illustrated in Figure 20-4.

20.5 Band Filters

Band filters may be obtained by connecting a low-pass and a high-pass filter in series, with their cutoff frequencies so arranged as to pass or suppress the desired band. In practice, however, these filters are designed and built as a single structure having two cutoff frequencies, using the same principles as already discussed for the low- and high-pass types. The band-pass filter shown in Figure 20-2 is one of several prototype forms. An m-type section can be derived for each of the various prototypes, and the one in question is illustrated in Figure 20-7. This m-type is derived in the same general manner as those for the low- and high-pass filters. In the band-pass filter, just as with low-pass and high-pass filters, it is necessary to use half T-sections for the end termina-

Composite Low · Pass Filter
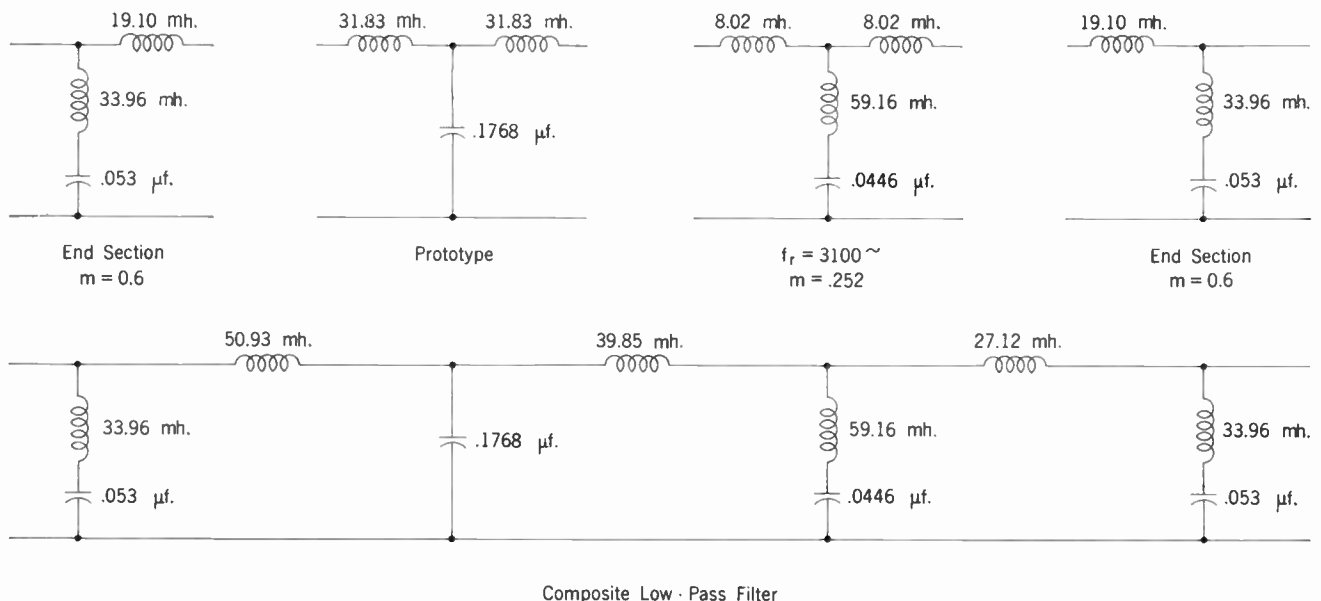
FIG. 20-6 MAKE-UP OF TYPICAL COMPOSITE LOW-PASS FILTER

[182]

Prototype

m · Type
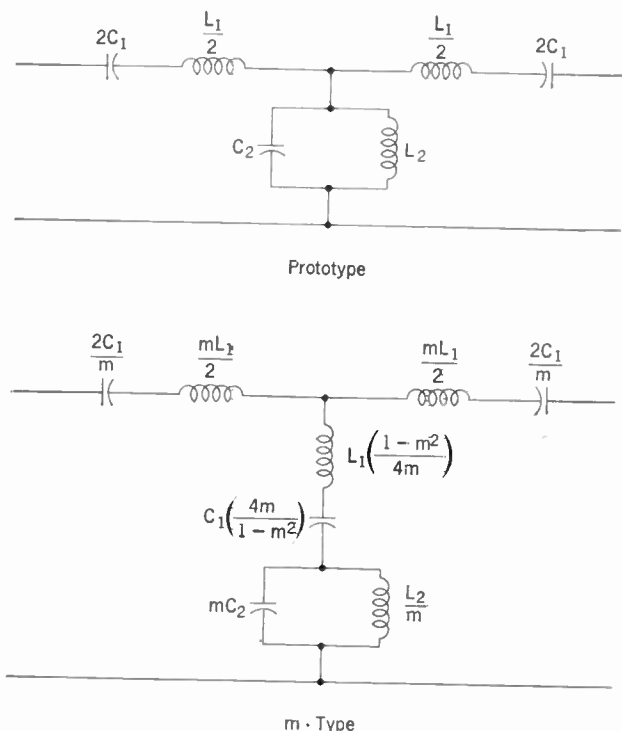
FIG. 20-7 BAND-PASS FILTER SECTIONS

tions with m equal to 0.6, thus providing uniform terminal impedances over the major portion of the pass band.

If f_l is the lower cutoff frequency, and f_u the upper cutoff frequency of the band-pass filter, the constants of the prototype are—

$$L_1 = \frac{R}{\pi(f_u - f_l)} \qquad (20:21)$$

$$C_1 = \frac{(f_u - f_l)}{4\pi f_l f_u R} \qquad (20:22)$$

$$L_2 = \frac{R(f_u - f_l)}{4\pi f_l f_u} \qquad (20:23)$$

$$C_2 = \frac{1}{\pi(f_u - f_l)R} \qquad (20:24)$$

where R is the characteristic impedance at zero angle of the circuit in which the filter is to be connected. The value of m for the m-type section is obtained from the following:

$$m = \sqrt{1 - \left[\frac{f_r(f_u - f_l)}{f_r^2 - f_u f_l}\right]^2} \qquad (20:25)$$

where f_r is the resonant frequency of infinite attenuation.

Since there are two cutoff frequencies, f_u and f_l, there should be two resonant frequencies to

provide infinite attenuation both above the upper cutoff frequency, f_u, and below the lower cutoff frequency, f_l. It may be noted, however, from an inspection of the m-type section of Figure 20-7, that for each value of m there will be two resonant frequencies in the overall network of the shunt arm. Therefore, when one resonant frequency, f_r, is selected, which determines the value of m, the other resonant frequency is also established. Both of these resonant frequencies, which result from one value of m, will occur at points having approximately the same percentage deviation from their respective cutoff frequencies. For example, in a band-pass filter designed to pass the frequency band between 20,000 and 23,000 cycles, if f_r is 2% above the upper cutoff frequency of 23,000 cycles, its value will be 23,460 cycles. By substituting these values in equation (20:25), the value of m will be found to be 0.627. This same m (0.627) also establishes an f_r for the lower cutoff frequency, which is 19,608 cycles or approximately 2% below 20,000 cycles.

20.6 Crystal Filters

In our discussion thus far, we have considered
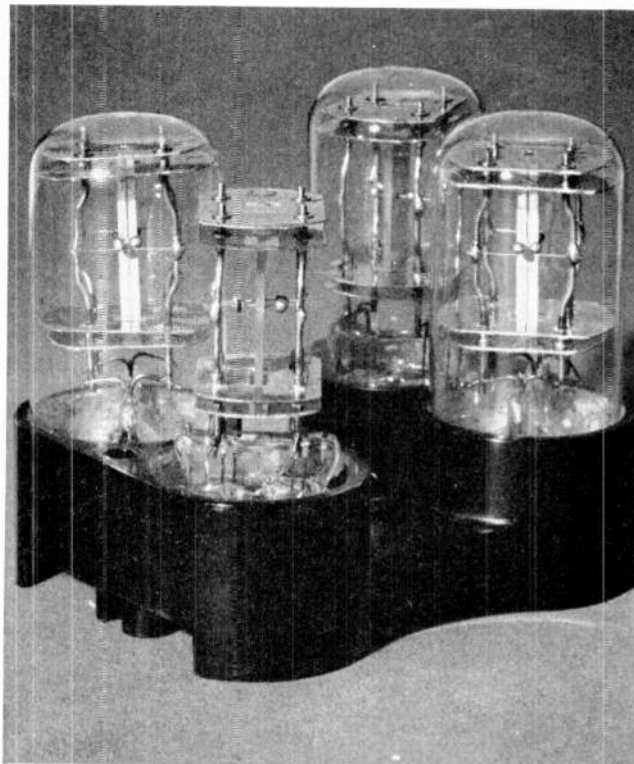
GLASS-SEALED CRYSTAL UNITS

filters built with ideal reactance elements having no resistance. This, of course, is not the case in actual practice. The presence of resistance in the inductors used in filter sections introduces additional losses in the transmitting bands, and reduces the sharpness with which the filter cuts off. In other words, the width of the transition band is increased by this resistance. In telephone and telegraph carrier systems, the number of channels which can be used in a given frequency range depends on the width of the pass band plus the transition bands on each side of it. Obviously, the shorter the transition bands, the greater the number of channels that can be obtained in a given frequency range.

Other things being equal, the sharpness with which a properly designed filter cuts off, or the width of the transition band, is determined by the ratio of the reactance to the resistance of the inductors used in the filter. As we have noted, this ratio of reactance to resistance is generally referred to as Q. The reactance of a coil of given inductance increases directly with an increase in frequency while its resistance remains fairly constant over the lower frequency range. At the higher frequencies, however, this resistance also increases with an increase in frequency, due to "skin effect" and other causes. As a result, the Q of a coil seldom exceeds 400 in practice. In the

EQUIVALENT ELECTRICAL CIRCUIT OF A QUARTZ CRYSTAL

REACTANCE CHARACTERISTIC OF A QUARTZ CRYSTAL

FIGURE 20-8

"T" FILTER SECTION USING CRYSTALS

REACTANCE CHARACTERISTIC CURVES

ATTENUATION - FREQUENCY CHARACTERISTIC
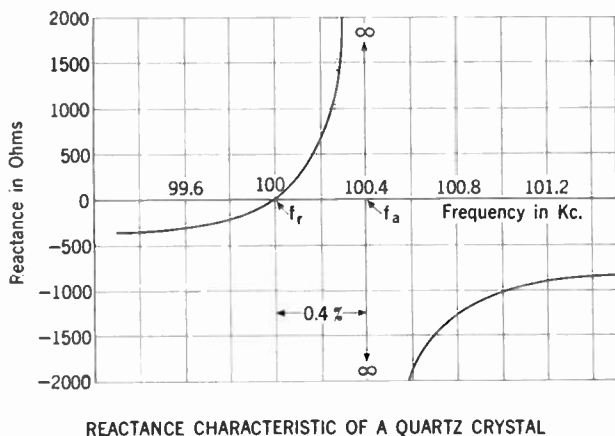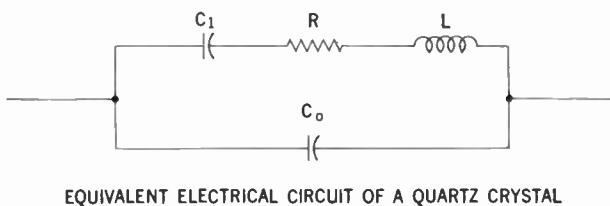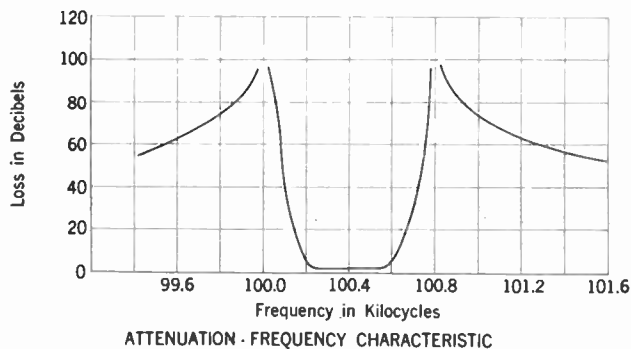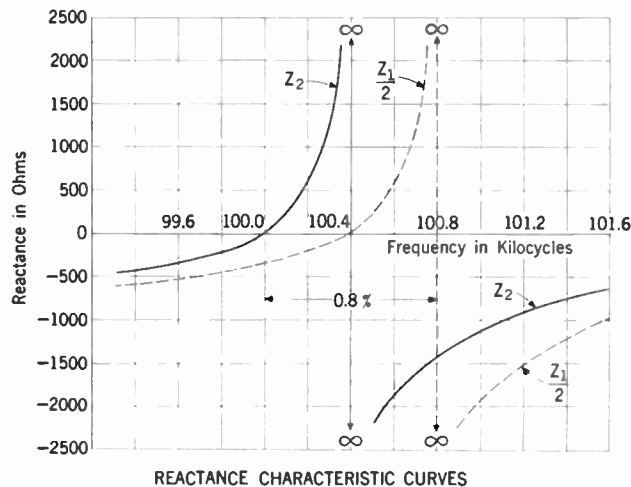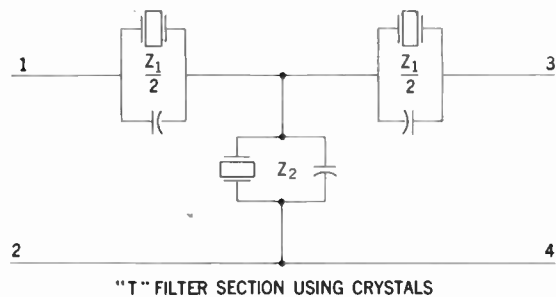
FIGURE 20-9

frequency range up to about 30 kc, this value of Q has been found high enough to provide satisfactory filters for carrier operation, but for frequencies above this value filter elements having higher Q's are desirable.

One of the most practicable ways to obtain such high Q elements is by the use of mechanical vibrating systems, such as the **piezo-electric crystal,** which possesses a natural mechanical-electrical relationship. Of the many substances that exhibit this piezo-electric effect, one that has been found most satisfactory for these purposes is crystalline quartz. Artificially grown crystals of certain chemical compounds are also used effectively.

[184]

Best known of these are EDT (Ethylene-Diamine-Tartrate) and DKT (Dipotassium-Tartrate). When a piezo-electric crystal is strained mechanically, an electric field is set up in its neighborhood, which will induce electric potential in any conductors in the field. Conversely, when a crystal is placed in an electric field, it is subject to mechanical strain and its shape changes very slightly. Thin slabs of the crystalline material when cut in certain ways will exhibit these same characteristics even more markedly. When the frequency of an applied alternating voltage is the same as the natural period of mechanical vibration of the crystal, the intensity of vibration of the crystal will reach a sharp "resonant" maximum. The natural frequency, or period of vibration, of the crystal depends on its dimensions, its density, and its elasticity.

In an electric circuit such as a filter, a crystal acts like an impedance that can be represented electrically as shown in Figure 20-8, where the inductance, L, represents the mass reaction of the crystal against motion (inertia); the resistance, R, represents the energy dissipating action in the crystal as it vibrates; C_0 represents the natural capacitance of the crystal when at rest (static characteristic); and C_1 represents the elasticity determining the storage of mechanical energy in the crystal (dynamic characteristic).

The ratio of C_0 to C_1 is a constant for any given crystal material. For quartz it is 125 to 1. The Q for quartz crystals is of the order of 20,000 or more. The general reactance characteristics of such a crystal are also shown in Figure 20-8 where the series resonant frequency is f_r, and the parallel resonant frequency is f_a. Since there is a fixed ratio of C_0 to C_1 (125 to 1) for quartz, the parallel resonant frequency, f_a, is always 0.4 per cent higher than the resonant frequency, f_r. This may be more easily understood if we keep in mind that resonance occurs when the mass reaction represented by L and the elastic reaction represented by C_1 are equal in magnitude but opposite in sign. For parallel-resonance, however, C_0 must be taken into consideration and the effective capacitance in the looped circuit becomes $\frac{125}{126}$ of that for the resonant condition, thereby making the parallel-resonant frequency higher by almost exactly 0.4 per cent. This means that the general form of the reactance characteristic of such a crystal is fixed. Of course the natural period of vibration of the crystal determines its position in the frequency scale, and this can be varied at will by cutting crystals of different dimensions.

If an auxiliary capacitor is placed in parallel with the crystal, however, the effective capacitance of C_0 is increased, and as a result the parallel-resonant frequency can be made less than 0.4 per cent above the series resonant frequency. This can be done without detracting from the favor-
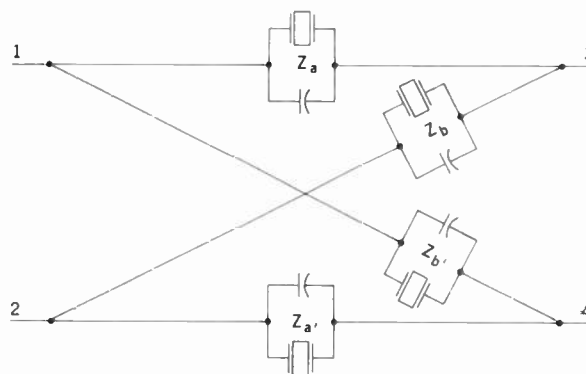
LATTICE FILTER SECTION USING CRYSTALS
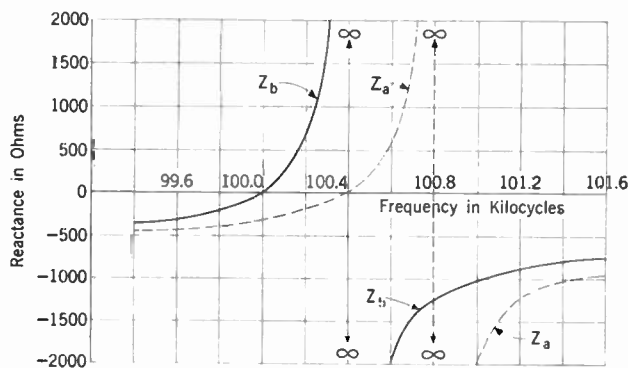
REACTANCE CHARACTERISTIC CURVES
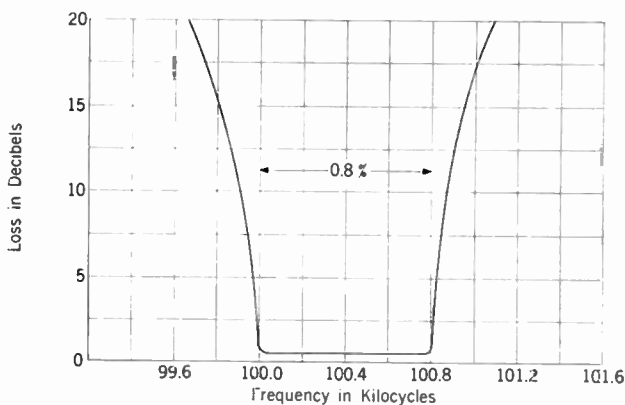
ATTENUATION - FREQUENCY CHARACTERISTIC

FIGURE 20-10

[185]
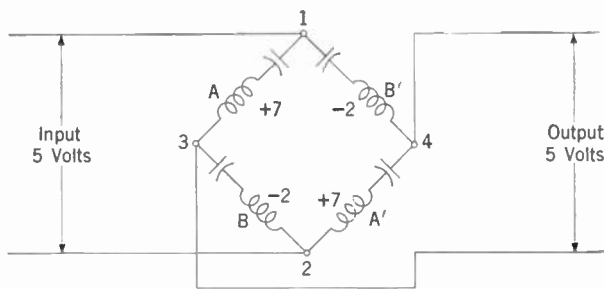
FIG. 20-11 PRINCIPLE OF LATTICE NETWORK

able characteristics of the crystal because the Q of the added capacitor will normally compare favorably with the Q of the crystal.

As we have seen, the simplest form of a filter circuit is the T-network. Now if crystals are placed in the series and shunt arms of such a network, we have the circuit shown in Figure 20-9, where capacitors are also added in parallel with the crystals to permit control of the band width. The characteristics of such a filter circuit can be analyzed by drawing the reactance curves for the series and shunt arms, as illustrated in this same Figure, where the auxiliary capacitors are assumed for the moment to have zero value. If the crystal elements are so selected that the resonant frequency of the series arm (zero reactance) coincides with the parallel-resonant frequency of the shunt arm (infinite reactance), the T-network will have points of maximum attenuation at the parallel-resonant frequency of the series arm (infinite reactance) and the series resonant frequency of the shunt arm (zero reactance). The pass band must therefore lie between these two points of maximum attenuation, as indicated in the lower drawing. Because of the limitation of the ratio of the frequency of parallel-resonance to that of series resonance, the peaks of maximum attenuation cannot be separated by more than 0.8 per cent of the frequency scale; and with the shunting capacitors, the separation will be less than this amount, depending upon the capacitor values.

20.7 Lattice Networks

A more general type of filter is obtained by using a bridge type (lattice) network as illustrated in Figure 20-10, where the two series arms are alike as

are the two shunt arms, but the series and shunt arms differ from each other. This type of network is equivalent to a bridge circuit where no current flows in the output when the bridge is balanced, which occurs at the frequencies when the reactances (impedances) of the shunt and series arms are equal.

To better understand this balanced condition, let us consider the circuit of Figure 20-10 where 1–2 and 3–4 are the input and output terminals, respectively. At a frequency where the series and shunt reactances are equal, there is the same voltage drop across 1–3 as across 1–4. This means that the voltage difference between 3 and 4 is zero, and the filter network has an infinite loss. When the frequency is such that the reactances in the series and shunt arms are of opposite sign, on the other hand, the voltage difference between the output terminals 3–4 is equal to the input voltage across 1–2, and current of this frequency is passed with zero attenuation (assuming no energy dissipated in the crystals).

This may be more easily seen from examination of this network rearranged in the more conventional balanced bridge form of Figure 20-11. Here the series arms, A and A', and the shunt arms, B and B', may be either positive (inductive) or negative (capacitive) reactance, depending upon the frequency of the input. For example, if an input of 5 volts produces a voltage drop across A and A' of +7 volts and across B and B' of −2 volts, there will be the same voltage in the output (3–4) as at the input. On the other hand,
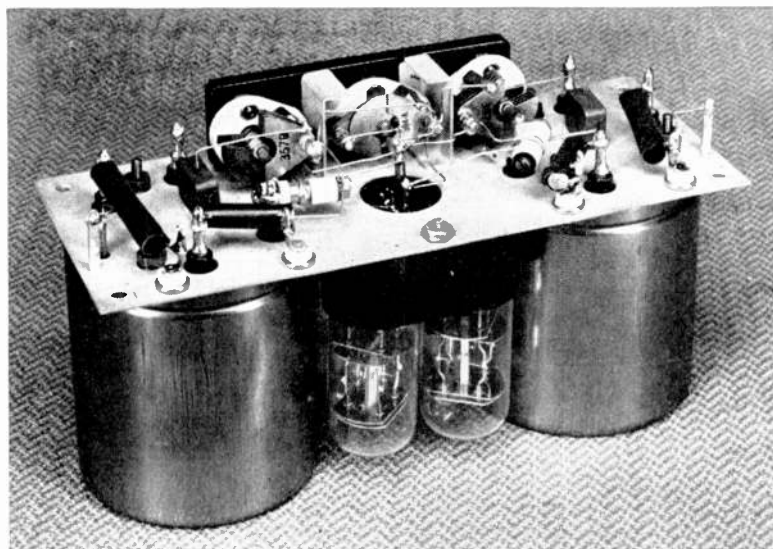
CRYSTAL CHANNEL FILTER
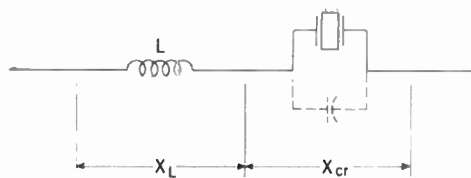
INDUCTANCE IN SERIES WITH CRYSTAL
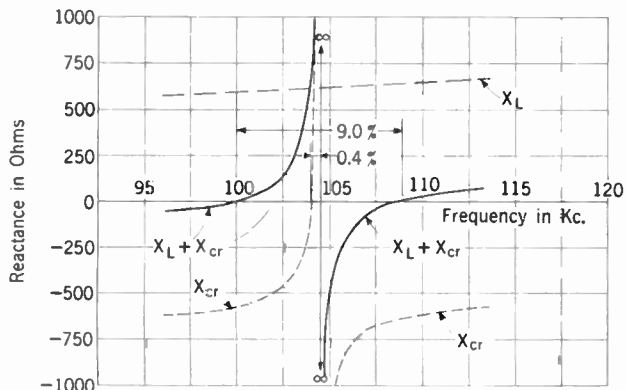
REACTANCE CHARACTERISTIC CURVES

FIGURE 20-12

(energy dissipated) they introduce does not over-come the beneficial effect of the low loss of the crystals. As a practical matter, it has been found that within certain limits the Q of the crystals has sufficient margin to make the use of such coils possible without serious detrimental effects.

LATTICE FILTER SECTION WITH SERIES INDUCTANCE

REACTANCE CHARACTERISTIC CURVES

ATTENUATION - FREQUENCY CHARACTERISTIC

FIGURE 20-13

when the frequency is such that the reactances in these two arms are of the same sign, the voltage across the output terminals 3–4 becomes less than that across the input. The loss of the network then takes some definite value between zero and infinity.

At both the resonant frequency for the shunt arm (zero reactance or short circuit) and the parallel-resonant frequency for the series arm (infinite reactance or open circuit), the voltage across the output terminals will be the same as that across the input terminals (assuming no energy dissipated in the network elements). These two frequencies are the cutoff values. This type of lattice filter network therefore passes the full 0.8 per cent band, and the pass band (and the corresponding peaks of infinite attenuation) may be placed in any position by varying the ratio of the impedances of the crystals in the series and shunt arms.

If we are working in the frequency range of 100 kc (100,000 cycles), however, the 0.8 per cent pass band is only 800 cycles, which is too narrow for a voice channel. In practice, wider pass bands are required. They may be obtained by adding inductance coils in the network of crystals. As we have seen such coils have a relatively low Q, and they will be satisfactory, therefore, only if they can be used in such a manner that the loss

[187]

Fig. 20-14 Typical Band-Pass Crystal Filter

If a coil is placed in series with a crystal, the reactance characteristics of the combination are as illustrated in Figure 20-12, where there are now two series resonant frequencies instead of one. This characteristic curve is obtained by adding the curve for the inductive reactance of the coil to the reactance curve for the crystal. The resonant frequencies occur at the two points where the combined reactance curve crosses the zero axis. It can be shown that when the parallel-resonant frequency is half-way between the two series resonant frequencies, the parallel-resonant frequency is 4.5 per cent greater than the first series resonant frequency. This means the two series resonant frequencies are now separated by 9.0 per cent.

By adding inductances to the lattice network of Figure 20-10 we have the filter illustrated in Figure 20-13, which passes the band of frequencies from the first resonant frequency of the series arm to the second resonant frequency of the shunt arm. The characteristic curves show that this filter arrangement now has a pass band of 3 x 4.5 or 13.5 per cent. In the frequency range of 100 kc this means a pass band of 13,500 cycles. The frequencies of infinite attenuation, of course, occur where the impedances of the two arms are equal

and of the same sign, which is another way of saying the bridge arrangement is balanced.

By varying the size of the capacitors in parallel with the crystals, the pass band may be made any width less than 13.5 per cent, although the loss caused by the series resistance in the inductors becomes rather large for widths of less than 0.5 per cent. Above 13.5 per cent width, the ordinary coil and capacitor filter is satisfactory, and below 0.5 per cent width, where the losses of the coils become serious, the all-crystal filter may be used. The impedance of the filter with the coil in series with the crystal is comparatively low—usually under 600 ohms—but by placing the coil in parallel with the crystal (which will have the same general effect on the width of the pass band as the series connection), this impedance may be made as high as 400,000 ohms for the narrower band widths.

Crystal filters have a wide field of application—especially where "broad-band" carrier is employed. A representative band-pass crystal filter, such as is used in the high-frequency cable carrier system (Type-K), is shown schematically in Figure 20-14. It may be noted that this filter contains two lattice sections. In each section, the crystal elements in the series arms are identical

[188]
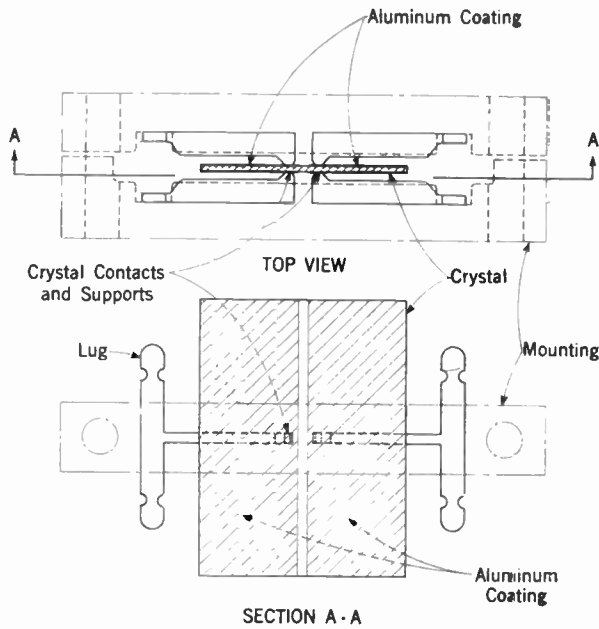
FIG. 20-15 METHOD OF MOUNTING QUARTZ CRYSTAL

range to obtain the same loss in each filter.

Electrically there are four crystal elements in each lattice section but for reasons of economy and for convenience in handling and adjusting the crystals, those in corresponding arms are "mechanically one". This means actually that there are but two physical crystals in each lattice section, or four for the whole filter shown in Figure 20-14. This is accomplished by plating the surfaces of the crystals used with a thin layer of aluminum which is divided along the center line lengthwise of the crystal to form two electrically independent crystal units from a single crystal. Since the crystals vibrate longitudinally with a node across the middle, they are clamped at this node in mounting as illustrated in Figure 20-15. The four crystals used in the lowest frequency channel (64 kc) of the Type-K carrier system range in length from about 1.65 to 1.58 inches, while those in the highest frequency channel (108 kc) range from about 0.96 to 0.94 inch. These

and those in the lattice (shunt) arms are identical. However, the elements of one section are not the same as those of the other section. The inductances are now cut in half and placed outside of the lattice. The shunting resistance between the two sections, together with the resistance components of the adjacent coils, form a resistance pad to match the impedance of the two lattice sections. By varying the size of this shunting resistance, the loss may be adjusted over a small

FIG. 20-17 TANK CIRCUIT

crystals vary in thickness from about 0.0248 inch to 0.0433 inch, while the widths are in the order of 50 to 80 per cent of their lengths.

Figure 20-16 is a representative loss-frequency characteristic of a two-lattice section crystal bandpass filter. The attenuation peaks nearest the pass
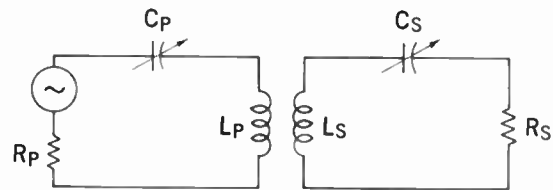
FIG. 20-18 COUPLED CIRCUIT

band result from the reactance of the series and lattice arms being equal in one lattice section. In a similar manner, the second set of attenuation peaks is due to the other lattice section. Considering cutoff as occurring at the points where the losses become 10 db greater than the loss at 1000 cycles, the frequency band passed is about 3600 cycles wide.

FIG. 20-16 LOSS-FREQUENCY CHARACTERISTIC OF FILTER SHOWN IN FIG. 20-14

20.8 Radio-Frequency Filters

At radio frequencies it is customary to speak of devices such as we have been discussing as selecting or tuning circuits, rather than as filters. And generally speaking, such devices at radio frequencies may be considerably simpler in design than many of the complex structures used in low frequency work. Thus the tuning circuits used in ordinary radio receivers are usually simple parallel-resonant **tank circuits** as indicated in Figure 20-17, which may be tuned to a particular frequency by varying the capacitance. In effect, such a circuit is a band-pass filter because, although in theory it is resonant at only one frequency, the resistance which is necessarily included in the circuit broadens the tuning so that appreciable bands of frequencies on each side of the resonant frequency are also selected. This is illustrated in the resonance curves shown in Figure 13-11. Percentagewise, the tuning may be quite sharp, but at radio frequencies a very small percentage deviation may well be sufficient to include a desired frequency band, such as that required for ordinary voice transmission. Thus, to pass a 10,000 cycle band where the radio carrier frequency is 1,000,000 cycles means a percentage deviation on either side of the carrier frequency of only 1%.

The selected frequency band may be broadened somewhat if necessary by using a transformer arrangement in which both the primary circuit and the secondary circuit may be separately tuned to resonance. One type of such an arrangement is shown in Figure 20-18. Such a circuit may be designed to serve the purpose of matching impedances as well as acting as an effective band-pass filter. Here if the primary and the secondary circuits are separately tuned to the same frequency, and the coupling between the primary and secondary windings of the transformer is adjusted so that the current in the secondary is at maximum (critical coupling) the inter-action between the two tuned circuits will be such that the frequency response of the secondary circuit will be broadened appreciably and the current will be
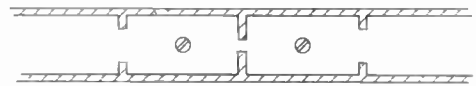
FIG. 20-20 TWO-CHAMBER FILTER IRIS-COUPLED

constant over a moderately wide band of frequencies. Closer coupling of the transformer windings will result in an even broader frequency response in the secondary, but at the expense of some decrease in the maximum current value.

20.9 Waveguide Filters

At super-high frequencies it is impracticable
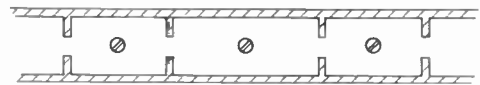
FIG. 20-21 CAVITY-COUPLED FILTER

to use tuned circuits, such as we have been discussing, employing coils and capacitors. Analogous principles apply, however, in the design of waveguide filters. Thus the resonant cavity, which we have already considered as being comparable in many respects to the parallel-resonant tank circuit, now becomes the basic element of the waveguide filter. Superficially, at least, if such a cavity is inserted in a waveguide it might be expected to pass a narrow band of frequencies centering about its resonant frequency, and to reject other frequencies.

Figure 20-19 (A) is a sketch of such a resonant cavity designed for insertion in a waveguide having the same cross-sectional internal dimensions as the cavity. The screw shown centered in the top wall of the cavity is known as a **trimmer** and is used for making fine adjustments of the reso-

A

B
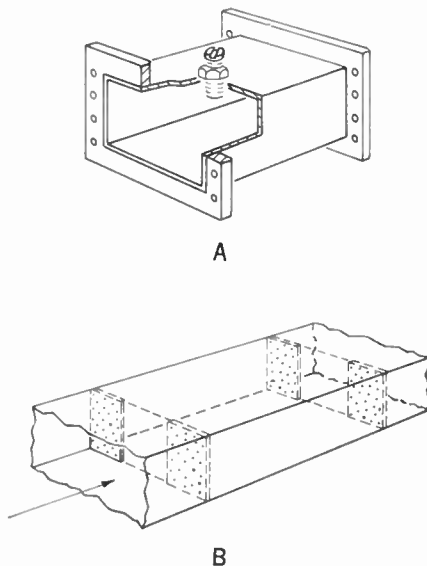
FIG. 20-19 WAVEGUIDE BAND-PASS FILTER SECTION

nant frequency. The screw end projecting into the cavity is located in the region of the strongest electric field. Its presence tends to shorten the electric lines of force and thus has an effect similar to increasing the capacitance in an ordinary tank circuit. In other words, inserting the screw farther into the cavity lowers the frequency of resonance of the cavity. This result is effectively the same as if the length of the resonant chamber were increased.

When the cavity is inserted in the guide its ends are enclosed by thin plates which, however, must contain openings to permit the wave of energy to pass through. The openings in the end plates are usually known as irises. In the arrangement shown in Figure 20-19 (B), the irises are vertical and the plate assembly has an effect like a shunt inductance because the plates projecting into the region of greatest magnetic field effectively shorten the lines of magnetic force. Plates, with their irises, of course represent discontinuities in the guide and will therefore tend to cause wave reflections and reflection losses. But, when the cavity section is of proper length (in the order of one-half wavelength) reflections returning into the waveguide line in either direction from the two discontinuities will be so phased as to cancel each other. At the particular frequency at which this is true, there will be a substantial standing wave within the cavity but from outside, the line will appear electrically smooth and power will be freely transmitted.

The Q of cavity resonators, as has been noted earlier, is very high as compared with that of any practicable resonant circuit made up of inductors and capacitors—in the order of 20,000 as compared to a few hundred. The resonance peak of the cavity is therefore relatively very sharp and for most purposes a single cavity inserted in a waveguide would not permit the passage of a sufficiently wide band of frequencies. This difficulty may be overcome by using two or more cavities in series. Such cavities could be tuned to slightly different frequencies but usually the tandem cavities will be tuned to the same frequency and coupled in such a way as to obtain an increased flat band of frequency transmission on the basis of the principles discussed in the preceding Article. Two types of coupled waveguide filters are in general use. The first, known as iris-coupled, where the resonant cavities are adjacent to each other and coupled by an iris having a considerably smaller opening than those at the

Band Width Adjusted By Top Screws

Frequency Adjusted By Side Screws
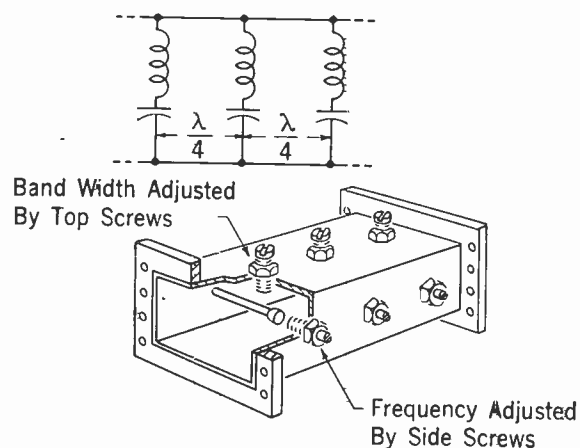
FIG. 20-22 BAND REFLECTION FILTER.

outer ends of the cavity, is illustrated by Figure 20-20. The second type, known as a cavity-coupled filter, is illustrated schematically in Figure 20-21. Here the two cavities, tuned for the desired frequency, are coupled by a third cavity which is tuned to a frequency well outside of the range of frequencies with which the particular situation is concerned. Waveguide filters may be built with practically any number of waveguide cavities in tandem, although each cavity causes some loss and it is therefore desirable not to use more than is necessary. In practice as few as three such cavities can be readily arranged to provide a flat pass-band of up to as much as 50 mc in width, with very sharp cutoff on each side.

Band-elimination filters as well as band-pass filters are frequently used in waveguide practice. These are also sometimes known as band-reflection filters. A simple filter structure of this type is shown in Figure 20-22 together with its analogous circuit equivalent. The transverse rod extending partially across the guide has an inductive effect, as it tends to shorten the lines of the magnetic field. The screw extending into the guide from the top has a capacitive effect, as we have already noted. Together they act like a series resonant shunt across the guide, which can be finely adjusted to present an effective "short-circuit" at a particular frequency. When three of these resonant shunts, spaced approximately a quarter-wavelength apart, are used as shown in the Figure, it is readily possible to adjust them so that the structure will reject quite sharply a band of frequencies as wide as 20 mc, while passing other frequencies freely. It should be noted that the energy of the rejected frequency band is

not absorbed in the filter but totally reflected. Furthermore, the design is such that the reflections from the several shunts add in phase. As we shall see in a later Chapter, this behavior is put to very practical use in the channel separating filter arrangement of microwave radio relay systems.

CHAPTER 21

ELECTRON TUBES

21.1 Vacuum Tube Elements

Of the several major types of electron tubes used in telephone and telegraph work, the best known and most widely employed is the high-vacuum thermionic tube, generally known simply as a vacuum tube, and familiar to users of radio receivers and comparable devices. This tube consists essentially of a metal or glass envelope from which practically all air or other gases have been evacuated, and into which are sealed two or more electrodes whose terminals may be connected to external circuits. One of these electrodes, known as the emitter or cathode, is the source of the free electrons that are essential to the tube's operation.

Electrons are supplied from the cathode by virtue of thermionic emission. This depends on the fact that certain metals or metallic oxides, when heated to a point of incandescence, tend to release electrons from their surface atoms into the surrounding space. Effectively, electrons "boil out" of the metallic surface in very considerable numbers to form a cloud of free electrons surrounding the electrode. This cloud is of course negatively charged and is known as the "space charge". If other electrodes in the tube are at the same potential as the cathode, the space charge cloud will quickly reach a stage of equilibrium depending on the temperature of the cathode, with electrons in equal numbers being emitted and returned to its surface. However, when a potential positive with respect to the cathode potential, is applied to another electrode, the electrons will be attracted to it and a continuous stream of electrons will flow from the cathode to the positively charged electrode, which is usually known as the anode or plate. This stream of electrons constitutes a flow of electric current (which, however, in accordance with the convention adopted long before electrons were discovered, is considered as flowing from the anode to the cathode).

Cathodes of thermionic vacuum tubes may be heated directly by a current of electricity supplied from a battery or other appropriate source. In this case, the cathode will be in the form of a fila-

ment as in an ordinary electric lamp bulb. Cathodes are frequently heated indirectly, in which case the cathode itself usually takes the form of a small oxide-coated cylinder, within which is enclosed an electrically insulated filament. This filament is also heated by an electric current but its electric circuit has no connection with the operating circuit of the tube. In most drawings of vacuum tube circuits, the heating element is for this reason not shown.

21.2 Two-Electrode Tubes—Diodes

Perhaps the simplest form of device for the flow of a stream of electrons through an open space is the two-electrode vacuum tube or **diode** illustrated in Figure 21-1. Here the cathode is indirectly heated from an external source and the plate battery B holds the plate at a positive potential with respect to the cathode. Under these conditions, as we have already noted for the general case, direct current will flow through the plate-cathode-B battery circuit and will register on the milliammeter. The value of current flowing for a given value of cathode temperature, and with other operating conditions remaining unchanged, will depend upon the voltage of the B battery. Figure 21-2 is a curve showing the relation between this space current and the plate voltage for three different values of cathode temperature. Here we find that as the plate voltage is increased from zero, there is an increase in the space current until a saturation point for the
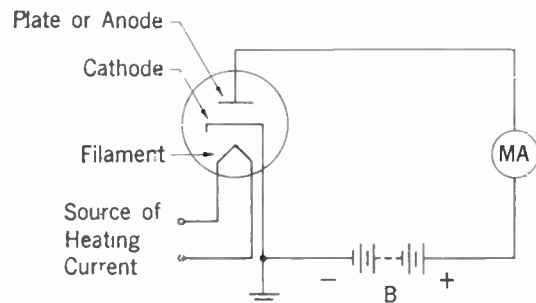
FIG. 21-1 TWO-ELECTRODE VACUUM TUBE

[193]
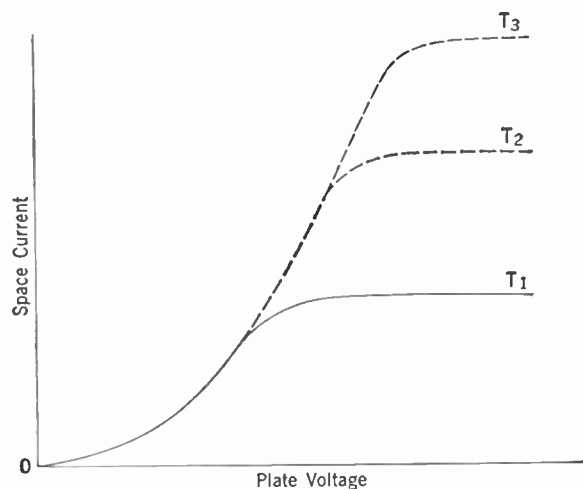
FIG. 21-2 PLATE VOLTAGE VS. PLATE CURRENT
CHARACTERISTICS

given temperature is reached. The failure of the space current to continue its increase further with increasing plate voltage is due to the fact that the cathode is emitting the maximum number of electrons possible for that particular temperature. If the cathode temperature is increased, the saturation point will increase correspondingly, as shown by T_2 and T_3 in the Figure. In practice, a vacuum tube is ordinarily operated at such a cathode temperature and plate voltage range that this saturation point will not be reached.

Under these normal operating conditions, the electron flow in the diode is dependent only on the plate voltage. Increases and decreases of plate voltage will cause corresponding increases and decreases of plate current. If the plate voltage is reduced to zero, or is made negative with respect to the cathode, there will be no current flow. The tube permits current flow in only one direction and, therefore, may be used as a rectifier or detector. Thus, if an alternating emf is substituted for the B battery in Figure 21-1, a space current varying in value for the positive half-cycle of voltage but completely cut off for the negative half of the cycle will result.

21.3 Three-Electrode Tubes—Triodes

The type of vacuum tube most widely employed in telephone work, differs from the two-electrode tube in that a third electrode or **control grid** is interposed between the cathode and the plate, as indicated in Figure 21-3. In this device, the electrons that leave the cathode must pass through the meshes of the control grid to reach the plate.

Their passage, therefore, is influenced by any force that may be set up by a charge on this control grid. Due to the relative positions of the control grid and plate with respect to the cathode, a change of potential of the control grid has a greater effect on the space current than an equal change in the potential of the plate. For example, a change of one volt in the potential of the control grid of a typical tube used in telephone repeaters, would have the same effect on the space current as a change of approximately 6.5 volts in the plate potential. The ratio of the change of voltage of the plate to the change in control grid voltage producing an equivalent effect, is called the **amplification factor** of the tube, and is usually designated by the symbol, μ. Its value depends entirely upon the mechanical design of the tube.

The utility of the triode tube in communication circuits is primarily due to the sensitive response in the plate circuit to small impressed potentials on the control grid. In this connection the control grid, in its control over the current in another circuit, is analogous to the valve of a water faucet. It decreases or increases the current in the plate circuit, and the force necessary to thus regulate it is independent of the value of the current or the amount of energy that may exist in the plate circuit. To best illustrate the relation between the control grid voltage and the current in the plate circuit, a curve is employed which is known as the characteristic operating curve or **transfer characteristic** of the particular type of tube. Figure 21-4 illustrates such a curve for a single value of plate voltage. Here the voltage that is impressed on the control grid, either positive or negative, is laid off to the right or left of the zero point, respectively, and the vertical ordinate shows the corresponding plate current value.

When the control grid is strongly negative with respect to the cathode, its field overpowers that
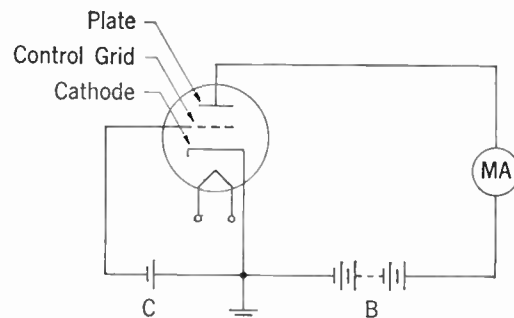
FIG. 21-3 THREE-ELECTRODE VACUUM TUBE

[194]
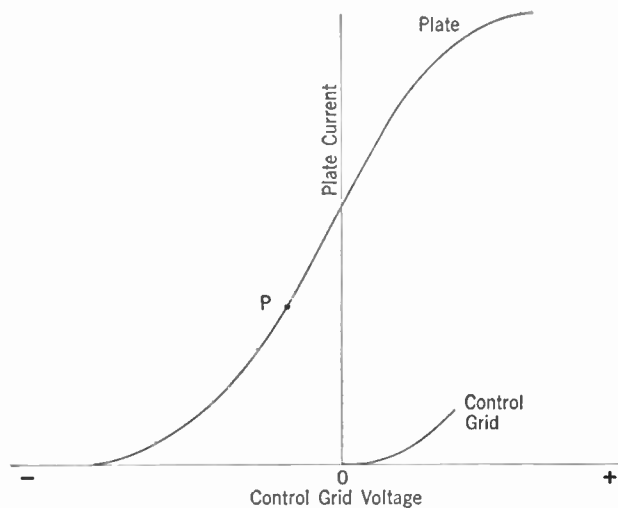
FIG. 21-4 CONTROL GRID VS. PLATE CURRENT
CHARACTERISTIC

due to the plate and electrons leaving the cathode
do not pass to the plate; i.e., the space current is
zero. If, now, the negative charge on the control
grid is gradually reduced, a point will be reached
at which the effect of the control grid no longer
overpowers that of the plate and a small current
flows. No current flows in the circuit of the con-
trol grid, which is still negative with reference
to the cathode and therefore repels the electrons.
The plate current rises according to the upper
curve of Figure 21-4. When the control grid be-
comes positive with respect to the cathode, it
draws some of the electrons to itself, establishing
a control grid current which varies as shown by
the lower curve so designated. The sum of the
control grid and plate currents is limited by the
ability of the cathode to emit electrons. Conse-
quently, as the control grid becomes more and
more positive, the plate current curve bends to-
wards a horizontal direction at its upper end and
may even fall again due to the control grid tak-
ing a larger share of electrons. The point at which
this flattening takes place depends on the tempera-
ture of the cathode, as pointed out in connection
with Figure 21-2.

At low frequencies, the circuit between the con-
trol grid and cathode is substantially open when
the control grid is negative since no current flows
through the space by transfer of electrons. As in
the case of any two metallic electrodes separated
by an insulating medium, however, there is some
capacitance in the cathode-control grid circuit,
which will permit some charging current to flow.
The capacitance is so small that the charging cur-

rent will be negligible at low frequencies, but at
radio frequencies this capacitance effect may be-
come important.

When the control grid has a negative potential
and the plate voltage is varied, the space current
curve differs from that shown in Figure 21-2 in
that the potential of the plate must be made great
enough to overcome the effect of the control grid
before any current will flow. This is illustrated by
Figure 21-5 for a single value of grid voltage. The
potential at which current begins to flow is μE_g,
where E_g is the voltage of the control grid. Above
this value the current varies as shown by the
curve of Figure 21-5, which is similar to the curve
for a two-electrode tube. Such a curve is known
as the **plate characteristic** of the tube.

In the design of triodes, one of the major char-
acteristics of the tube to be considered is the am-
plification factor μ. As noted earlier, this is the
ratio of a small change in plate voltage to a cor-
responding small change in grid voltage, with the
plate current held constant. This may be repre-
sented by the equation—

$$\mu = -\frac{E_{p2} - E_{p1}}{E_{g2} - E_{g1}} \qquad (21:1)$$

In more mathematical language, μ is the slope at
a given plate current value of a curve in which
grid voltage is plotted against plate voltage. Its
value may therefore be written as—

$$\mu = -\frac{\Delta E_p}{\Delta E_g} \qquad (21:2)$$

where Δ indicates a very small change in the
quantity which it precedes. The minus sign is
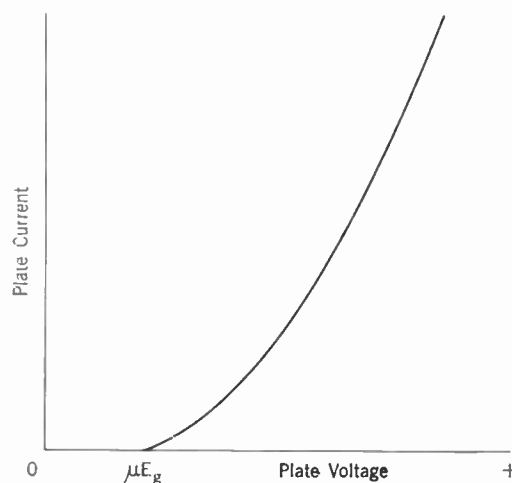
FIG. 21-5 PLATE VOLTAGE VS. PLATE CURRENT
CHARACTERISTIC OF THREE-ELECTRODE TUBE

[195]
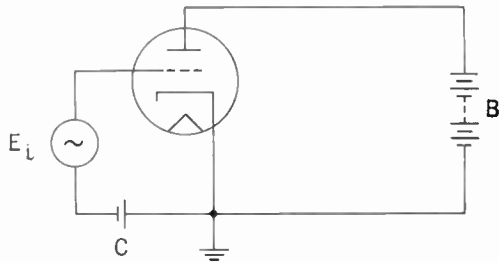
FIG. 21-6 SIMPLE TRIODE CIRCUIT

necessary because the simultaneous changes in plate voltage and grid voltage are in opposite directions.

Another fundamentally important characteristic of a vacuum tube is its output or **plate resistance**, which is usually designated R_p. This is somewhat different from a resistance in the ordinary sense because the ratio between plate voltage and plate current is not a simple application of Ohm's Law. Plate resistance of the tube is more in the nature of an "a-c resistance". It is the ratio of a small change in plate voltage to a corresponding small change in plate current, with the grid voltage held constant. In other words, it is the slope of a plate voltage-plate current characteristic curve such as is shown in Figure 21-5, or—

$$R_p = \frac{\Delta E_p}{\Delta I_p} \qquad (21:3)$$

A third basic characteristic of a vacuum tube is its **transconductance**, usually designated G_m. This is defined as the ratio of a small change in plate current to a corresponding small change in grid voltage, with the plate voltage held constant. It is the slope of the grid voltage-plate current characteristic curve of Figure 21-4. That is—

$$G_m = \frac{\Delta I_p}{\Delta E_g} \qquad (21:4)$$

where G_m will be in mhos when I_p and E_g are in amperes and volts respectively.

Inspection of equations (21:2), (21:3) and (21:4) will show that the three characteristics are inter-related as follows—

$$\mu = G_m R_p; \ R_p = \frac{\mu}{G_m}; \text{ and } G_m = \frac{\mu}{R_p} \qquad (21:5)$$

21.4 Dynamic Characteristics of Tubes

Curves of the types shown in Figures 21-4 and 21-5 are known as "static" characteristic curves

because they are plotted on the basis of fixed d-c potentials at all points. In nearly all practical applications of vacuum tubes, a-c or rapidly varying potentials are superimposed on the fixed d-c potential. Thus in Figure 21-6, the alternating voltage E_i is shown impressed on the grid in series with the fixed negative potential of the battery C. The outstanding characteristic of this simple circuit is its ability to amplify small voltage changes impressed by E_i to comparatively large current changes in the plate circuit. The grid voltage-plate current characteristic of the tube, which is of the form shown by Figure 21-4, is redrawn in Figure 21-7. Here the alternating input voltage E_i is shown as a sine wave impressed on the characteristic curve at the point P, where the distance from P to the vertical zero axis indicates the fixed negative bias maintained by the grid battery. The resulting plate current is shown as a sine wave about the horizontal axis drawn to the right from P. The form of the plate current wave is a faithful reproduction of the a-c wave impressed on the grid as long as the portion of the tube characteristic involved is a straight line. Furthermore, as long as the varying grid voltage remains negative with respect to the cathode, there is no current flowing to the grid and no power is expended in the input of the circuit.

The circuit of Figure 21-6 is interesting but it obviously has little practical value unless some means is employed to obtain useful output voltage or power. One method of doing this is to connect a load resistance, R_L, in the plate circuit as shown in Figure 21-8. In this case, current flowing in
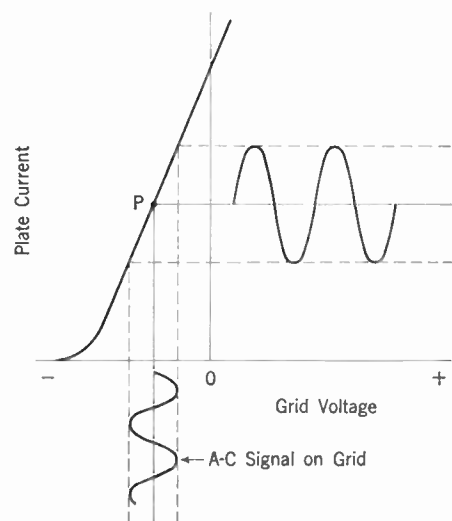
FIG. 21-7 VACUUM TUBE AS AN AMPLIFIER
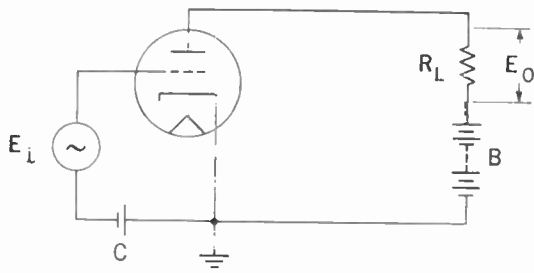
FIG. 21-8 VACUUM TUBE WITH LOAD RESISTANCE

the plate circuit will cause a proportional voltage drop across the load resistance R_L. The output voltage E_o, represented by this voltage drop can then be applied to other circuits. The presence of the load resistance, however, will change the shape of the grid voltage-plate current characteristic of the tube. Instead of the static characteristic curve of Figure 21-4, the new curve will be a "dynamic" characteristic. This is illustrated in Figure 21-9, where A reproduces the static characteristic curve of Figure 21-4, and B is the dynamic characteristic resulting from the addition of the load resistance R_L. The difference is due to the fact that the voltage drop across R_L, being in series with the B battery, causes the positive potential at the plate to vary with the amount of current flowing in the plate circuit in accordance with the following—

$$E_p = E_B - I_p R_L \qquad (21:6)$$

where E_B is the d-c voltage applied in the plate circuit.

In Figure 21-9, P is the point on curve A corresponding to the fixed value of grid bias produced by the C battery. In order that the dynamic characteristic may pass through this point, it is assumed that the value of the B battery is increased until the net voltage at the plate is the same as in the static condition with no load resistance present. P is thus established as a common reference point for the two curves. When the grid voltage increases, the resultant increased plate current causes an increased voltage drop across R_L. This reduces the plate potential and thus prevents the plate current from increasing to the same extent that it would increase without the load for a like increase in grid voltage. This means that the dynamic characteristic falls below the static characteristic at all points to the right of P. On the other hand, when the grid voltage decreases below the value at P, the decrease in plate potential due to the voltage drop in the load

resistance is less than the amount by which the B battery voltage was increased. The dynamic curve, therefore, lies above the static curve at all points to the left of P. The net result, as indicated in Figure 21-9, is that the dynamic curve is considerably flattened and is more nearly a straight line than the static curve. The difference in slope also shows that the amplification of the circuit with an output load may be expected to be less than that indicated by the amplification factor of the tube.

The voltage amplification of the circuit of Figure 21-8—i.e., the ratio of the output voltage E_o to the input voltage E_i, is readily calculated from Figure 21-10. Here the tube and its input has been replaced by an equivalent circuit. This, in accordance with Thevenin's Theorem, consists of the tube output voltage, which is equal to μE_i, in series with its output resistance R_p, as shown at the left of points 1 and 2. From this simple circuit it is evident that—

$$E_o = \frac{\mu E_i R_L}{R_L + R_p}$$

and from this, the voltage amplification of the circuit is—

$$\frac{E_o}{E_i} = \frac{\mu R_L}{R_L + R_p} \qquad (21:7)$$

Inspection of this last equation shows that the voltage amplification or "gain" of the circuit is always less than the amplification factor of the tube, but that it approaches the amplification factor as the load resistance increases.
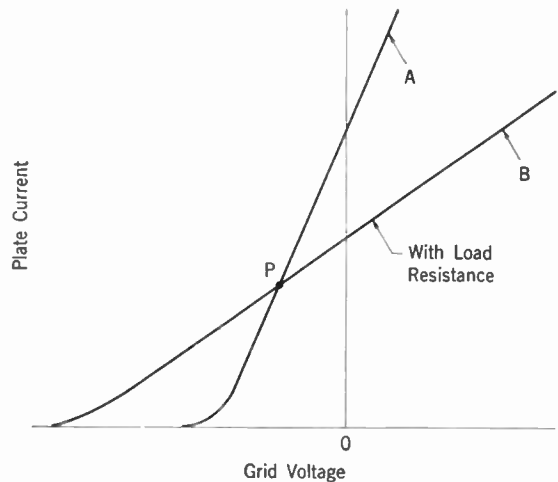
FIG. 21-9 EFFECT OF LOAD RESISTANCE ON PLATE CURRENT-GRID VOLTAGE CHARACTERISTIC CURVE
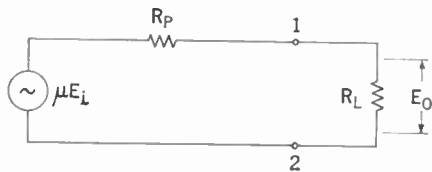
FIG. 21-10 EQUIVALENT CIRCUIT OF FIG. 21-8

The voltage amplification can also be determined graphically from a static plate characteristic curve by the method illustrated in Figure 21-11. Here is plotted a family of plate characteristic curves for different values of grid bias voltage. Across these curves a straight line known as a **load line** is drawn. This load line applies to only one specific load resistance value. It merely represents the effect of the voltage drop in the load on the plate potential, and is exactly defined by equation 21:6 ($E_p = E_b - I_p R_L$). The intersection of the load line with the horizontal axis where $I_p = 0$ is obviously at the point where $E_p = E_b$. The intersection with the vertical axis, when $E_p = 0$, would be at the point where $I_p = E_b/R_L$. The intersection of the load line with any one of the plate voltage-plate current characteristic curves represents the zero input condition of the tube for the particular value of grid bias applying to that curve. When an alternating voltage is applied to the grid, the instantaneous relation between plate current and plate voltage will then be represented by a point swinging back and forth along the load line through the points of intersection.

The application will perhaps be better understood by means of a concrete example. The characteristic curves of Figure 21-11 are drawn for a representative triode connected in a simple am-
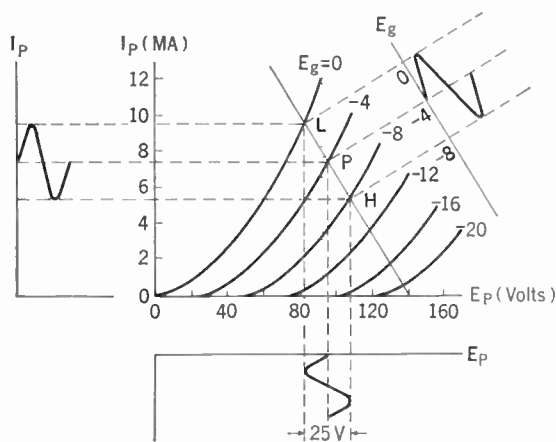
FIG. 21-11 GRAPHICAL ANALYSIS OF TUBE OPERATING CHARACTERISTICS

plifying circuit as in Figure 21-8. The plate supply voltage E_b is assumed at 140 volts, grid biasing voltage -4 volts, and the load line is drawn for a load resistance of 6000 ohms. The static condition of the tube is represented by the intersection of the load line with the -4 volt curve at point P as indicated. The values of E_p and I_p are then 95 volts and 7.5 milliamperes respectively. The corresponding voltage drop across the load resistance is then 45 volts (which is equal to $140 - 95$ and to $.0075 \times 6000$). Now, if an alternating voltage having a maximum swing of plus or minus 4 volts is applied to the grid, the operating point of the tube will swing along the load line between points (L) where it intersects the $E_g = 0$ curve, and (H) where it intersects the $E_g = -8$ curve. These points establish the maximum and minimum values of the plate voltage as 107 and 82 volts respectively; and the maximum and minimum values of the plate current as 9.6 and 5.4 milliamperes respectively. Thus the diagram shows that an input peak-to-peak voltage variation of 8 volts on the grid results in a variation of 25 volts at the plate. The voltage amplification under these conditions of operation is, therefore, 25/8 or 3.1.
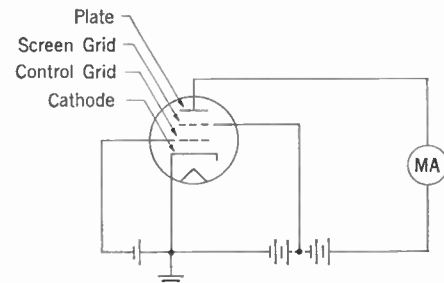
FIG. 21-12 FOUR-ELECTRODE VACUUM TUBE

21.5 Multi-Electrode Tubes

There are many designs of vacuum tubes containing more electrodes than the tubes we have been considering. Most widely used of these are four electrode tubes or **tetrodes** and five electrode tubes or **pentodes**. The basic theory of operation of such tubes is essentially the same as that of the triode. The additional electrodes act to improve the operating characteristics with respect to the amount of amplification to be obtained and may have other desirable effects.

At relatively low frequencies, the amplification factor μ of a triode can be made to have almost

any desired value by properly spacing and proportioning the three electrodes. When tubes are used with higher frequencies such as are encountered in radio and other high-frequency systems, the effect of inter-electrode capacitance becomes increasingly important. This is particularly true of the capacitance between the plate and control grid, where its coupling effect may be especially troublesome because it provides a path between the input and output of the tube through which output energy may feed back into the input circuit. This plate-control grid capacitance effect can be practically eliminated by placing a shielding grid between the control grid and the plate as illustrated in Figure 21-12. This grid is known as a **screen grid,** and the four-electrode tube is then known as a **screen grid tetrode.**

As the plate is shielded by the screen grid from the other electrodes, it (plate) has little effect in withdrawing electrons from the space charge area about the cathode. This function is taken over by the screen grid which is given a positive potential for this purpose. The flow of electrons from the cathode, and their control by the control grid, is practically the same as discussed in the
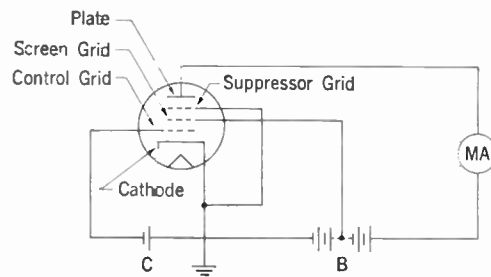
Fig. 21-14 Five-electrode Vacuum Tube

case of the three-electrode tube, but in the screen grid tube, the screen grid itself may be considered as acting in somewhat the same manner as did the plate of the three-electrode tube. However, the electrons constituting the space current, on arriving in the area of the screen grid have acquired such a velocity that most of them pass through the openings in the screen grid and, attracted by the still higher positive voltage of the plate, continue on to the plate. A small portion of the electrons is, of course, intercepted by the screen grid and does not reach the plate. This is illustrated by the plate voltage vs. plate and screen grid current curves in Figure 21-13 for a representative screen-grid tetrode.

In the normal working range of the tube, where the characteristic curve is relatively flat, it will be noted that the plate current change is quite small for a considerable change in plate voltage. This means that the output resistance of the tube is very high as compared to the triode. Due to the presence of the screen grid, the variation of plate voltage has relatively little effect on the plate current, but the control grid retains the same control of plate current as in the triode. The amplification factor is accordingly much higher.

It will be noted, however, that at plate voltages close to or less than the fixed screen grid voltage, the characteristic curves show a pronounced drop in the plate current. This is due in part to the fact that the screen under these conditions is drawing an excessive part of the cathode current because of its relatively high positive potential. More important is the fact that it is now attracting electrons emitted by the plate as a result of **secondary emission.** This emission is caused by the high-speed electrons striking the plate with such force as to knock some of the outer electrons out of the plate material. Under normal operating conditions these secondary electrons will fall back into the plate due to the influence of its positive
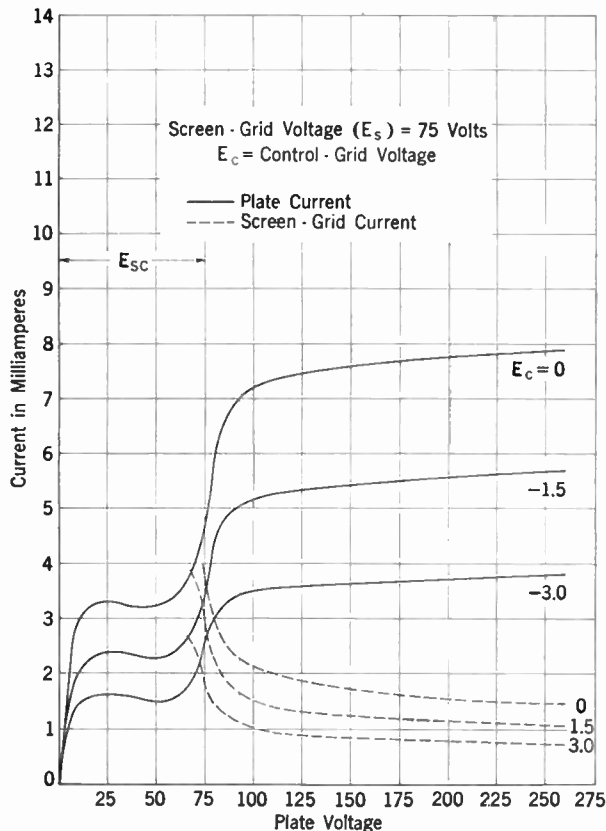
Fig. 21-13 Characteristics of Four-electrode Tube

[199]
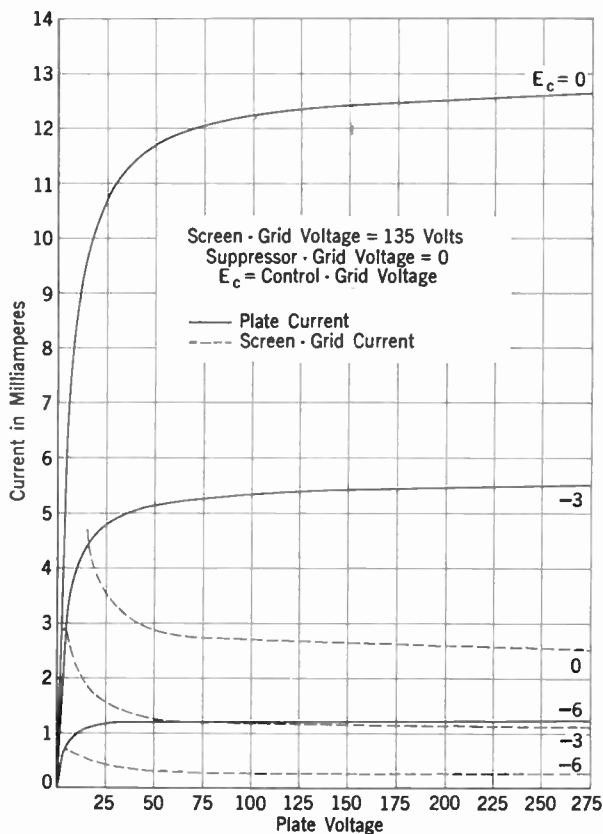
FIG. 21-15 CHARACTERISTICS OF FIVE-ELECTRODE TUBE

potential; but if the screen potential is as high or higher than the plate potential, some of them will be attracted to the screen, thus effectively reducing the total flow to the plate and increasing the screen-cathode current. Obviously, operation of the tube in this region would result in marked distortion of the output signal. Generally, therefore, screen grid tubes must be operated with a plate supply voltage sufficiently high so that the maximum negative swing resulting from the control grid input signal will not reduce the instantaneous plate potential to a value approaching that of the screen.

While there is always some secondary emission of electrons from the plate of a tube, its effect on the tube's operating characteristics can be practically eliminated by introducing another grid between the plate and the screen grid, as shown in Figure 21-14. This grid, which is maintained at a potential negative with respect to the plate, is called a "suppressor grid" and the tube then becomes a **suppressor-grid pentode**. The suppressor grid is usually connected directly to the cathode, often inside the tube. Its field repels the secondary electrons emitted from the plate, forcing

them back to the plate. Figure 21-15 shows characteristic curves for a tube of this type. It will be noted that like the screen grid tetrode, both the amplification factor and output resistance are high.

Typically, the small pentode designed primarily for voltage amplification has a plate resistance R_p of more than 1 megohm, with an amplification factor μ somewhere between 100 and 5000. The extremely high plate resistance tends to limit the usefulness of such a tube for power amplification. More effective for the latter purpose is the **beam power tube** which is fundamentally a tetrode that secures the effect of the pentode's suppressor grid by increasing the electron space-charge in the region between the screen grid and the plate.

The structure of this tube is indicated by the diagram of Figure 21-16, which represents a lateral cross-section. As shown, the cathode is flattened with the control and screen grids surrounding it at relatively close spacing. Both elliptical grids are wound with the same number of turns and these are so aligned that each wire of the control grid is directly in front of the screen grid as viewed laterally from the cathode. The spacing between the screen and the plate is relatively large. The beam-forming plates at each side are connected directly to the cathode. When the tube is in operation, these plates act to focus the
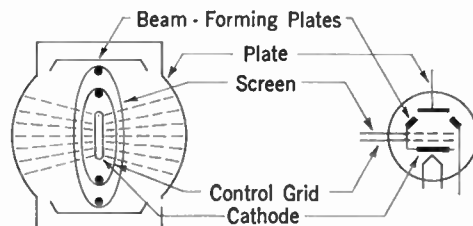
FIG. 21-16 BEAM POWER TUBE

electrons emitted by the cathode into lateral beams as indicated in the drawing. The concentration of the electrons, together with the wide spacing between screen and plate, results in relatively large numbers of electrons being present in this space at all times. The negative space charge thus established in the screen-plate space is large enough to hold the potential within the space to a value lower than the voltage of either the plate or the screen. It accordingly acts to suppress secondary emission from either of these electrodes in much the same way as the suppressor grid of the pentode does. The characteristics of the tube are

[200]

therefore very similar to those of the pentode. However, the lateral alignment of the wires of the control and screen grids effectively shields the screen from the electron beam so that there is practically no screen current. The amplification factor is comparable to that of the pentode, although generally somewhat lower. The plate resistance is substantially lower (in the order of 50,000-100,000 ohms) which accounts for the advantage of the beam tube for power amplification.

There are many other possible designs of multi-electrode tubes, some practical types of which contain as many as eight electrodes. It is customary also to employ multi-unit tubes in which two or more independent electronic circuits are included in a single envelope. The structure of such tubes is indicated by their designations such, for example, as duplex-diode, duplex-diode-triode, twin-pentode, etc.

CHAPTER 22

ELECTRON TUBES—(*Continued*)

22.1 Gas Tubes

Gas tubes are electron tubes whose envelopes have been evacuated, and then partially refilled with a small carefully regulated amount of an inert gas, such as argon or neon, or a metallic vapor, usually of mercury. Gas tubes are commonly built as diodes or triodes, and sometimes as tetrodes. They are generally similar in appearance to high-vacuum tubes but have considerably different electrical characteristics.

In the thermionic gas diode, which is symbolically represented as shown in Figure 22-1(A), where the black dot indicates the presence of gas in the tube envelope, electrons will be emitted by the heated cathode just as in the case of the high-vacuum tube. Also, if a relatively positive potential is applied to the plate, the electrons will flow from the cathode to the plate. As the plate voltage is increased, the plate current will tend to build up along a characteristic curve similar to that of the high-vacuum tube. But, as the number and velocity of the moving electrons increases, they will collide in increasing numbers with the gas atoms in the intervening space. Some of these collisions will cause ionization of the atoms—i.e., separation of an outer electron from the atom, leaving a positively charged ion. Such action increases the number of electrons moving toward the plate, and further adds to the net current by creating a flow of positive ions moving toward the cathode. The net result is that after a certain critical potential difference is reached, the tube "breaks down" or effectively "arcs over" and becomes a relatively good conductor. The critical voltage at which this occurs is known as the **ionization potential** or **firing point** of the tube. After ionization has occurred, the value of the current

flowing is determined by the resistance in the external circuit, and it may increase up to the maximum permitted by the emission capabilities of the cathode without any further increase in the plate potential. As a matter of fact, current will continue to flow freely even when the plate potential is reduced considerably below the firing point. There is a minimum sustaining value of plate potential, however, below which the gas will deionize and conduction will stop.

FIG. 22-2 PLATE CURRENT-PLATE VOLTAGE CHARACTERISTIC CURVE OF A GAS-FILLED DIODE

The behavior of the tube is illustrated by the heavy curve of Figure 22-2, which also shows by the dotted curve the comparable behavior of a high-vacuum tube. It may be seen that the tube is effectively an electronic switch, which closes at a certain critical voltage to permit a relatively large current to flow, and opens at some lower voltage to stop the current flow. When the voltage is below the firing point, the tube has an extremely high resistance, but as soon as ionization occurs, its resistance is very low. Thus, it is a switch that

A   B

FIG. 22-1 GAS-FILLED DIODE AND TRIODE TUBES

[202]

can be closed almost instantaneously, and with practically no power, by the mere application of a suitable voltage to the plate. On opening, however, an appreciable time is required for the gas to de-ionize completely. This time will ordinarily be in the order of a hundred microseconds.

The gas triode is constructed with a third electrode or grid placed between the cathode and anode in much the same way as the control grid of a high-vacuum tube, as shown in Figure 22-1 (B). The operation of the tube is not fundamentally different from the gas diode, except that the firing point may now be controlled by the voltage applied to the grid; and since the grid is closer to the cathode than to the plate, the voltage to be applied to start ionization is smaller. The gas triode cannot act as an amplifier, however, because once ionization is started, the grid has no further control of the current flowing across the tube, unless it is made so negative as to cause de-ionization.

In both gas diodes and triodes employing heated cathodes, the construction is such that the anode is much larger than the cathode. Because of this fact, and the fact that the relatively heavy positive ions naturally travel at a much lower speed than the electrons, the flow of current across the tube is essentially in one direction only. These tubes may, therefore, be used as rectifiers, as discussed briefly in a later Chapter.

Gas tubes in which the cathodes are not heated are known as "cold-cathode" tubes. In this case, there are no electrons "boiling out" of the cathode but there are always a few free electrons floating about in the cathode-anode space. If a relatively positive voltage is applied to the anode of such a tube, these free electrons will be attracted towards the plate. As their velocity is increased, they will ionize atoms with which they collide. If the applied voltage becomes sufficiently high, the tube will break down and conduct a current in the same way as does the hot-cathode tube. As would be expected, the break-down voltage of the cold-cathode tube must be higher; and the total cur-
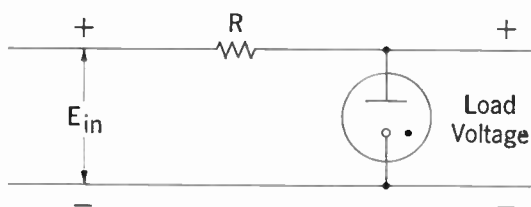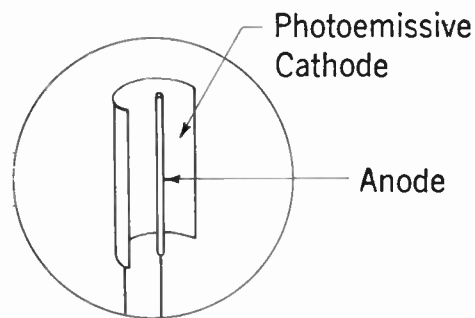
FIG. 22-4 SCHEMATIC OF PHOTOTUBE

rent that can flow in the space, following ionization, is lower than in the hot-cathode tube because the cathode is not supplying electrons by thermionic action. Within these limitations, however, the cold-cathode tube has the same general characteristics as the heated tube, and has the advantage that it does not require a continuous source of power for its operation. It is, therefore, particularly useful as a switching device where the required operation is highly intermittent. In cold-cathode tubes, the cathode is ordinarily much larger than the anode, with the result that the flow of electricity in the ionized tube may be in both directions. The tube, therefore, has no application as a rectifier.

An important feature of all gas tubes is that after ionization, the voltage drop across the tube (cathode-to-anode) is constant over a considerable range of current variation. This is a characteristic of the arc discharge itself, and has nothing to do with the positioning or shape of the tube elements. The total voltage drop occurs in a thin film surrounding the cathode, and its value depends entirely upon the nature of the gas with which the tube is filled. This characteristic leads to the extensive use of gas tubes as voltage regulators. Thus, if a tube is placed across a circuit as shown in Figure 22-3, and the applied voltage is sufficient to start ionization, the load voltage will take a value depending only on the tube voltage drop characteristic, and will remain constant at that value regardless of variations in the applied voltage.

22.2 Phototubes

The phototube is ordinarily a high-vacuum tube having two electrodes. Its action depends upon the property of certain metals, notably sodium, potassium, rubidium, and caesium, to emit elec-

FIG. 22-3 SIMPLE VOLTAGE-REGULATOR CIRCUIT

trons when irradiated with visible or ultra-violet light. As indicated in Figure 22-4, the photo-emissive cathode is much larger than the anode. The electrons which are emitted as a result of light falling on the photo-active metal of the cathode, are drawn away to the positively charged anode, thus establishing a small current of electricity which varies in strength with the intensity of the light.

In communications work, phototubes are used principally in connection with telephotography. Here they perform the essential function of producing an electric current which varies with the intensity of a beam of light reflected from the picture being transmitted. A much larger field of use for these tubes is in sound motion picture projection where they translate the "sound-track" on the edge of the film into sound. They are also employed extensively for operating relays under light control in various industrial applications.

22.3 The Cathode-Ray Tube

The cathode-ray tube is a highly specialized type of electron tube which has many uses as a testing and measuring instrument, although its current widest application is as the "picture tube" of the ordinary television receiver. Tubes of this type were in use in many electrical laboratories long before the invention of the tubes discussed in the preceding Articles.

Figure 22-5 shows the general structure of a representative "electro-static" cathode-ray tube. The cathode, which is indirectly heated, is a small cylinder, oxide-coated at the end so that it has the ability to emit electrons in large quantities. The control electrode (sometimes called the grid because of its analogy to the corresponding electrode of the ordinary triode) is a metal cylinder with a circular hole in its forward end, which is
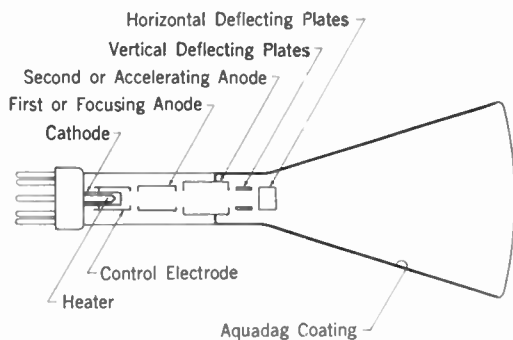
FIG. 22-5 ELECTROSTATIC CATHODE-RAY TUBE

located close to and partially surrounding the cathode. The control electrode serves the double purpose of controlling the number of electrons emitted by the cathode, and of focusing the emitted electrons so that they will form into a narrow beam emerging through the hole in its end. As in other tubes, the controlling effect of this electrode is affected by the application to it of a voltage negative with respect to the cathode.

In front of the control electrode is an anode, known as the first or focusing anode, which is maintained at a potential positive with respect to the cathode (usually by several hundred volts). This electrode is also in the form of a cylinder, somewhat larger in diameter than the control electrode, having holes in its ends through which the emitted electrons may pass. A second anode, known as the accelerating anode, maintained at a still higher voltage (usually one thousand or more volts) is placed in front of the first anode. This is a still larger cylinder likewise having small holes in the ends. The two anodes act together to attract electrons from the cathode, and to cause their forward movement at high velocity (up to as much as 25,000 miles per second). Some of the emitted electrons will of course strike one of the anodes and be neutralized, but because of their high velocity, and because the arrangement of the two anodes acts as a sort of "lens" to focus the electrons into a narrow beam, most of the emitted electrons will proceed ahead through the holes in the several electrodes to the distant end of the tube. The four electrodes together are commonly called an **electron gun.**

The electrons reaching the distant end of the tube, which is coated on the inside with a fluorescent material, give up their energy of motion causing the fluorescent screen to emit light from the point of impact. The spent electrons then find their way to the conducting coating of aquadag with which the sides of the tube are lined, and are returned to the accelerating anode. Incidentally, it may be noted that the tube is usually operated with the accelerating anode at ground potential, which of course means that the potential of the cathode is highly negative.

The practical usefulness of a cathode-ray tube depends upon deflecting the beam of electrons from the gun in such a way that a pattern will be traced on the screen. In the electrostatic tube shown in Figure 22-5, these deflections of the beam are caused by voltages applied to the two pairs of deflecting plates. Thus, if one of the

CATHODE-RAY TUBES

the geometry of the tube that voltages of the same frequency applied simultaneously in phase to the two pairs of plates will produce a straight diagonal line; while the same two voltages applied 90° out of phase will produce a circle.

More significant as a practical matter, is the fact that the tube may be used to display the pattern of a voltage wave of practically any form or frequency. For this purpose, the voltage being investigated is applied to the vertical deflecting plates. At the same time there is applied to the horizontal deflecting plates a sweep voltage having a saw-tooth pattern in which the potential difference increases at a constant rate over a time equal to the period of the voltage applied to the vertical plates, and then returns quickly to its starting point. The pattern of the wave applied to the vertical deflecting plates will then be repeated continuously so that it will appear as a steady picture on the screen.

For the picture tube of a television receiver, both sets of deflecting plates are connected in sweep circuits to move the beam over the surface of the screen in synchronism with the scanning devices of the "television camera". The control electrode varies the intensity of the electron beam (and the light spot) in accordance with the received television picture signal.

Cathode-ray tubes may also employ electromagnetic means for both focusing and deflecting the electron beam. An electromagnetic tube is shown in Figure 22-6. It will be noted that both the focusing and deflecting coils are outside of the tube envelope, which helps to make the tube somewhat more rugged than the electrostatic tube. The focusing coil is in the form of a ring, which fits over the neck of the tube. Its ability to focus the electron stream from the gun into a narrow beam depends on its location with respect to the anode and the amount of current flowing in its

vertical deflecting plates is made positive with respect to the other, the negative beam of electrons, as it passes between them, will be deflected in the direction of the more positive plate, and the point of light on the screen accordingly will move vertically. Similarly, a difference of potential on the horizontal deflecting plates will cause the light spot to be deflected horizontally. Because of the almost negligible mass of the electrons in the beam, there is practically no inertia, and the beam can be moved about with extreme rapidity. Thus, if a high frequency sine wave voltage is applied to the horizontal deflection plates, the spot of light will sweep back and forth in a horizontal line at a rate equal to the frequency of the applied voltage. Due to the persistency of phosphorescence of the screen material, a horizontal line of light will be seen on the screen. The same voltage applied only to the vertical deflecting plates will similarly produce a vertical line on the screen. It follows from

FIG. 22-6 ELECTROMAGNETIC CATHODE-RAY TUBE
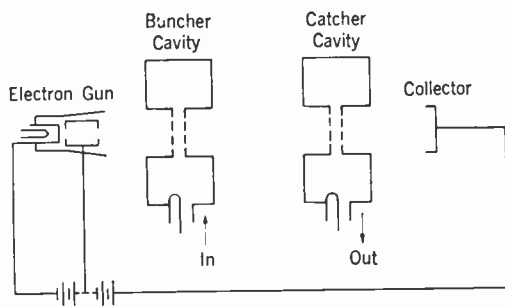
FIG. 22-7 BASIC STRUCTURE OF A KLYSTRON

windings. The deflection coils are in two pairs, as indicated. The electron beam is of course effectively a current of electricity and like all other electric currents is accompanied by a magnetic field. This field inter-acts with the fields of the deflection coils when they are energized in such a way as to exert pressure on the electron beam, just as pressure is exerted on a current-carrying wire in the presence of an external magnetic field.

In addition to greater ruggedness, the electromagnetic tube has the advantage that large deflections of the electron beam may be obtained without the employment of the excessively high voltages that would be necessary for like deflections in an electrostatic tube of equal total length. Tubes of this type are, therefore, generally used in television receivers where relatively large pictures are required. The electrostatic type has its chief application in laboratory devices such as the cathode-ray oscilloscope, where it is somewhat more effective than the electromagnetic tube when working with very high frequencies.

22.4 The Klystron

In the operation of electron tubes, it is sometimes necessary to give consideration to the time required for the electrons to travel between cathode and plate. In the usual triode, this "transit time" is normally in the order of 10^{-8} to 10^{-10} seconds, depending on electrode spacing and applied voltage. This is an exceedingly brief interval, but if an alternating voltage as high as 4000 or 5000 mc is applied to the grid of a triode having such a transit time, the polarity of the voltage on the grid will have time to vary from one-half to several cycles while an electron is moving across the inter-electrode space. It is obvious that the behavior of the tube in such circumstances would be seriously affected.

While triodes are built with electrode spacing

sufficiently close to overcome the adverse effect of transit time at super-high frequencies (up to approximately 5000 mc), there are entirely different designs of electron tubes which make use of transit time in their operation, instead of allowing it to be a limitation. One such tube is the klystron, which is also sometimes referred to as a velocity-variation or velocity-modulation tube.

The principle of operation of the klystron may be followed by referring to Figure 22-7. Included within the evacuated envelope (not shown in the Figure) is an electron gun, two sets of closely spaced grids, and a collector plate. Each of the sets of grids forms part of a resonant cavity, which is tuned to the frequency at which the tube is to operate. The electron gun is generally similar to the gun of the cathode-ray tube described in the preceding Article. It develops a beam of high velocity electrons which travel in a straight line through the grids of the two cavities to the collector plate.

Now, assume that the buncher-cavity is energized at the super-high frequency to which it is tuned by virtue of its size and shape. Electrons passing through its grids during a negative part of the frequency cycle will be slowed down slightly, and electrons passing during the positive half-cycle will be speeded up slightly. As they proceed onward through the so-called "drift space" between the buncher and catcher grids, they will tend to form into bunches—as the speeded up electrons overtake those that were slowed down. One such bunch tends to form for each cycle of the buncher-cavity oscillation. By the time the electrons reach the grids of the catcher-cavity, accordingly, they are traveling in bunches rather than in their normal random spacing.

This electron stream constitutes a current which is effectively varying in amplitude at the frequency of the buncher-cavity. The varying current, in passing through the grids of the catcher-cavity, induces a varying voltage between the two grids just as a varying current in the plate circuit of an ordinary vacuum tube will induce a corresponding varying voltage in a tuned tank circuit to which the plate circuit is coupled. If the catcher-cavity is tuned to the same frequency as the buncher-cavity, the bunched or modulated beam of electrons will transfer energy to the catcher-cavity, which can be delivered to the load as useful output power.

As an amplifier, the klystron tube can be designed to give a power gain of as much as 30 db

and to handle a frequency band several megacycles in width without serious distortion. These tubes are also capable of delivering very high power output, as much as 10 kw or more in some designs. In general, however, they introduce too much noise to permit their extensive application as amplifiers of small microwave signals.

22.5 The Traveling-wave Tube

For amplifying signals at frequencies higher than about 5000 mc, an entirely different structure, known as a **traveling-wave tube**, has many advantages. This device depends upon the interaction between the field of a beam of moving electrons and the field of a signal wave traveling at approximately the same speed in an evacuated tube. There are various possible structures that will encourage such interaction but the helix arrangement illustrated in Figure 22-8 is the one that seems generally most useful.

As there indicated, this device employs a long evacuated tube equipped with an electron gun at one end and a collector at the other. The electron gun is generally similar to that used in a klystron or a cathode-ray tube. In a typical case, a positive potential of about 1500 volts with respect to the cathode is applied to the anode and the collector. With the cathode properly heated, this produces a beam of electrons traveling from the cathode to the collector at a speed approximately one-thirteenth that of the speed of light. The beam is focused narrowly in the center of the tube by a powerful axial magnetic field of several hundred gauss intensity, which is produced by magnets surrounding the tube (not shown in Figure 22-8).

Within the long section of the tube, the electron beam passes through a loosely wound helix of wire which is connected to the signal input and output lines by means of coaxial or waveguide couplers. Signal current flowing through the helix wire sets up an electric field that is directed along its central axis and occupies the same position as the electron beam. Current in the helix wire of course

Heater — Anode   Helix   Collector

Cathode   Input   Attenuator   Output

FIG. 22-8 TRAVELING WAVE TUBE

flows at the speed of light but the velocity of the axial electric field which it produces is a fraction of this depending on the diameter and the number of turns of the helix per unit length along the axis. Successful operation of the tube requires that the diameter and the number of turns in the helix be such as to make the velocity of the axial electric field closely approximate that of the electron beam. In a typical tube designed for operation in the 4000 mc range, the helix has about five turns per signal wavelength. The total helix length is about four wavelengths or approximately twelve inches. Tubes designed for operation in other frequency ranges would have dimensions scaled proportionately, with all dimensions decreasing as frequency of operation is increased.

When the velocities of the electron beam and the signal wave are approximately the same but with the electrons moving at the slightly higher speed, a traveling-wave tube such as described above may give a power gain of 30 db or more. It is also capable of delivering output power measured in tens of watts. Furthermore, since it includes no tuned cavities or other resonant circuits, it can handle a very broad band of signal frequencies without distortion. Uniform amplification of signal bandwidths ranging up to 1000 mc is readily attainable.

Mathematical analysis of what takes place in the interaction of the traveling signal wave and the electron beam is somewhat complex but a qualitative physical picture may be developed from the following considerations. The moving electric field of the signal, which occupies the same space geometrically as the electron beam, of course changes its polarity cyclically in accordance with the signal frequency. As it cycles from negative to positive, it tends to increase the velocity of the beam electrons somewhat; as it changes in the other direction, it tends similarly to retard the beam electrons. Each cycle of the signal wave thus causes some bunching of the beam electrons. Since the beam electrons and the signal wave are traveling at essentially the same speed, they are effectively stationary with respect to each other and the bunching effect occurs at regular spaced intervals equal to one wavelength of the signal wave. The magnitude of the effect is cumulative so that the bunches grow progressively larger as the wave travels along the tube. Considering only the beam current, the bunching effect then appears as a superimposed wave which
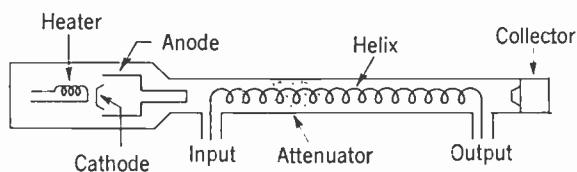
has the same frequency as the input signal wave and which increases in strength progressively along the tube. Since it has the same frequency, this new wave is removed with the original signal wave at the output terminal. The total output, which is the sum of these two waves, may thus be much greater in magnitude than the original signal wave but differs from it otherwise only by a small shift in phase.

Like practically all other amplifying devices, the traveling-wave tube is potentially capable of developing spontaneous oscillations. Relatively small impedance mismatches at the input and output terminals may cause reflected waves to travel back and forth in the tube and perhaps build up oscillatory energy values that may interfere seri-ously with the desired operation. In practice, this hazard is guarded against by inserting an attenuator at a point near the tube input as indicated in Figure 22-8. Usually this is effected by painting a short section of either the inside or outside of the glass tube with a lossy material such as aquadag. Energy reflected back from the tube output is effectively suppressed by this attenuator. Of course, it will also attenuate the forward traveling input wave but if the attenuator is placed far enough beyond the input to permit bunching of the beam electrons to start, the bunching effect will continue to take place through the lossy section and the magnitude of the bunching wave will increase steadily in the relatively long section remaining.

CHAPTER 23

SEMI-CONDUCTOR PRINCIPLES AND DEVICES

23.1 Conduction Theory

The conduction of electricity through a vacuum, as in the ordinary electron tube, is readily explained as the physical transition of electrons from one electrode to another under the influence of an appropriately applied electric field. It is true that no one actually sees the electrons in motion but the indirect experimental evidence of their presence and behavior is so voluminous as to leave no real doubt as to what is going on. It is customary, too, to explain electric current in a metallic conductor, such as a copper wire, as a stream of electrons flowing within the conducting metal. This theory is perhaps not as strongly supported by experimental evidence but it appears to account successfully for most of the observed phenomena of ordinary electrical conduction.

The theory assumes that there are always present in a good conductor, such as copper, very large numbers of free electrons which have become detached by thermal agitation from the outermost orbit shells of the individual atoms. At ordinary temperatures, the number of such free electrons in copper at any instant is thought to be about equal to the total number of atoms, which is 8×10^{22} per cubic centimeter. In the absence of any electric field, thermal energy causes the electrons to move about in rapid random motions, colliding with each other and with the atoms, whose movements are relatively quite limited. Many return to complete a deficient atom shell in the process; as many others are being simultaneously detached from other atoms to leave these atoms positively charged or "ionized" for the moment. The net result of all this activity, from the viewpoint of an external observer, is exactly nothing. However, if a difference of potential is externally applied to the two ends of the conductor, the resultant electric field impels the free electrons to move together in the direction of the field. Individually, the rapid random movements of the electrons continue but the application of the electric field results in a net mass movement of electrons through the conductor at a relatively slow rate. Because the free electrons are so very numerous, this adds up statistically to the passage of a vast number of electric charges per second through any cross-section of the conductor. This, by definition, constitutes an electric current.

It is an oversimplification, however, to assume that the conducting ability of all materials may be determined solely by the degree of freedom with which the outermost electrons of the individual atoms may become detached. Experimental evidence indicates that conductivity is also affected by the atomic or molecular structure of the material—i.e., the way in which the individual atoms or molecules are associated with each other. Thus, certain metals, such as lead and aluminum, when cooled to a temperature approaching absolute zero ($-459.6°$F.) exhibit a phenomenon called superconductivity. A current induced in a closed circuit of such a material will continue to flow almost indefinitely without further application of external force. A reasonable theory would appear to be that the absence of heat causes the individual atoms to "freeze" in fixed positions, leaving unimpeded passage for the free flow of electrons through the conductor. But other conductors, such as copper, which normally contain even more free electrons than the metals mentioned above, do not display superconductivity when cooled in the same way.

It is necessary to say that a completely satisfactory theory to explain the behavior of superconductors has not yet been developed although the phenomenon is currently a subject of intensive study and bids fair to have future practical applications in electrical communication and allied fields. Of even more current interest and application are the semi-conductors, of which germanium, silicon, selenium and certain metallic oxides are representative. Devices employing such materials are already in extensive use in both telephone and telegraph practice.

Explanation of the behavior of the semi-conductors must also be regarded as in a somewhat theoretical stage although a considerable volume of literature on the subject has grown up in recent years—especially following the invention of the

transistor in 1948. The following Articles attempt to present the theory now most generally accepted in relatively simple physical terms with minimum recourse to quantum theory, wave mechanics, and the associated specialized mathematics which is necessary to a more rigorous analysis.

23.2 Semi-Conductors

Of the semi-conducting materials currently in general use in electrical communications practice the metallic element, germanium, has perhaps been most thoroughly studied. Techniques have been developed for refining it to an almost incredible degree of purity and for processing it in the form of single crystals which can be treated or "doped" with impurity substances in any desired proportion. Unless specifically stated otherwise, the following discussion will be confined to this material. The principles developed will apply in general to all other semi-conductors, although their specific characteristics may of course differ.

Germanium is a metal which in ordinary electrical practice would be considered a very poor conductor. At room temperatures, typical "n-type" germanium, such as might be used in a transistor, contains one free electron for about 10^8 atoms as compared with one free electron for each atom in a good conductor such as copper. Even so, a cubic centimeter of this material includes something in the order of 10^{11} free electrons so that it is certainly not a good insulator. To understand its capabilities and limitations as a conductor, it is necessary to examine the germanium atoms and their structural arrangement in detail.

Chemically, germanium has atomic number 32 and valence 4. This means, in terms of modern atomic theory, that the atom nucleus is surrounded by four shells of orbital electrons containing 2, 8, 18, and 4 electrons. The first three shells are
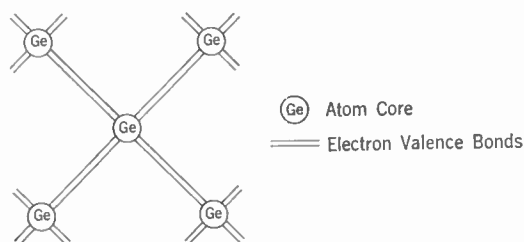
FIG. 23-1 VALENCE BOND STRUCTURE
OF GERMANIUM

completely filled and so firmly bound to the nucleus that they can play no part in any practical conduction process. The outermost shell, with its four valence electrons, is much less than half filled. Considering an individual atom alone, theory would predict that one or more of these four electrons might be readily broken away by thermal agitation or other relatively small forces. In the structure of germanium, however, the individual atoms are associated with each other in a crystalline lattice arrangement such that there is a definite bond between each atom and four immediately adjacent atoms. The structure is of course three-dimensional but the principle may be represented by the two-dimensional drawing of Figure 23-1. Two valence electrons—one from each atom—act together to form each bond, as repre-
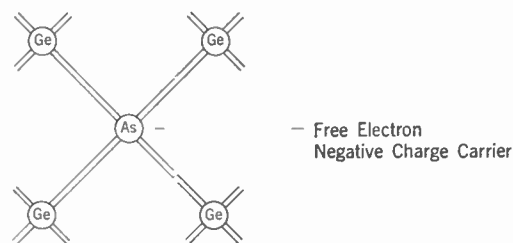
FIG. 23-2 N-TYPE GERMANIUM STRUCTURE

sented by the two connecting lines between the atoms. In this configuration, the valence electrons are held in place much more strongly than they are by the attraction of the atom nucleus alone. In germanium, it requires some fourteen times as much energy to free an electron from a valence bond as it would to detach it from an isolated atom.

Nevertheless, some electrons are knocked out of their valence bonds by thermal agitation at normal room temperatures. These electrons are free to act as "charge carriers" and the electrical conductivity of the germanium depends upon their number. At higher temperatures, more electrons are broken from valence bonds and the conductivity of the semi-conductor increases proportionately. This relationship of conductivity to temperature readily accounts for the negative temperature-resistance characteristics of the semi-conductor devices known as thermistors, which are widely used in telephone work for automatic transmission regulation and other purposes.

Thermal agitation at normal temperatures does not provide enough charge carriers to make pure

[210]

germanium capable of satisfactory transistor action. It is necessary for this purpose to modify the semi-conducting material so as to increase the number of carriers in roughly controllable amounts. This is done in two ways. In the first, pure germanium is "doped" with a small amount of an element with valence 5, such as arsenic (atomic number 33). Each arsenic atom will enter into the lattice structure in the manner indicated in Figure 23-2. Four of its outer-shell electrons join in valence bonds with four neighboring germanium atoms. The fifth valence electron is still associated with the arsenic atom but is quite easily dislodged. The net effect is to increase the number of free electrons in the material in proportion to the relative amount of arsenic added. In practical applications, this ratio is in the order of one impurity (arsenic) atom to 10 million germanium atoms. The semi-conducting material is now called **n-type** germanium because it contains a more than normal number of mobile negative charges.

Another way to increase the number of carriers is to dope the germanium with an element having valence 3, such as gallium (atomic number 31). As illustrated in Figure 23-3, the gallium atom also takes a position in the lattice structure. But its outer shell provides enough electrons to enter into valence bonds with only three of the germanium atoms, leaving one bond incomplete. The missing electron in the valence bond may be considered as a **hole** in the structure, which has an effective positive charge exactly equal but opposite to that of an electron. The hole may be filled by an electron which has been dislodged from another nearby valence bond by thermal agitation or other means. This process leaves a hole in the other bond so that the hole may be considered as having moved from one point to another in the material. Thus, in net effect, the material now contains a substantial number of effective positive charges that are relatively free

+ Hole
Positive Charge Carrier
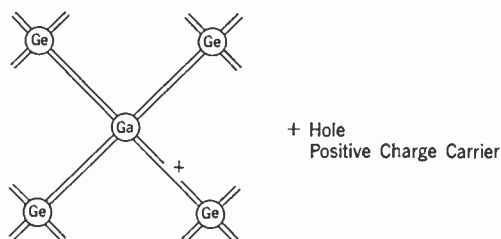
FIG. 23-3 P-TYPE GERMANIUM STRUCTURE

to move about in much the same way as free electrons may do. Such a material is called **p-type** because its mobile charge carriers are positive.

At first approach, the concept of holes as positive charge carriers may be regarded as a somewhat fanciful way of describing the transition of electrons from one valence bond to another. Actually, the electron in a valence bond is in a different "quantum state" than a valence electron in its ordinary position in the outer atomic shell of an isolated atom. In order to satisfactorily explain various observed conduction phenomena, it is not only convenient but necessary to treat the hole left by the ejection of an electron from a valence bond as if it were a real particle having mass, positive charge, energy, and velocity. This concept is reinforced by the observed fact that the mobility of positive and negative carriers is different. In germanium, electrons move, under a given electric potential, at somewhat more than twice the speed of holes (about 3600 cm per sec per volt per cm for electrons; about 1700 for holes).

Whenever an electron escapes from the valence shell of an atom, the atom is left with a net positive charge. Such atoms are called **donors** and are present in relatively large numbers in n-type material. Similarly, when a hole escapes from a trivalent atom, the atom has a net negative charge because an extra electron has been added. These atoms, called **acceptors**, are common in p-type material. Both donors and acceptors are therefore ions (i.e., electrically charged atoms), but they are locked in position in the lattice structure of the material and can play no direct part in any conduction process. The conducting ability of the semi-conductor is thus due solely to the presence of free electrons and free holes, and is proportional to the sum of all such carriers—both positive and negative. In pure, or intrinsic, germanium at normal temperatures, the number of free electrons is equal to the number of holes because whenever an electron is released from a valence bond by thermal agitation, a hole is simultaneously created. Holes and electrons will of course be continuously neutralizing each other through new recombinations but a like number will be escaping from other valence bonds at the same time so that the net conductivity remains constant under static conditions. In n-type germanium, there are many more free electrons than holes. These are then designated the "majority carriers" but there will always be some holes
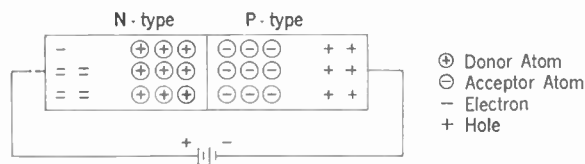
FIG. 23-4 P-N JUNCTION—REVERSE
CONNECTION

present, too, as "minority carriers". The opposite
situation prevails in p-type material, with holes
now taking the role of majority carriers. Finally,
it may be noted that the intrinsic material (pure
germanium) will display certain n-type character-
istics because, even though the number of holes
and electrons is the same, the mobility of the
electrons is appreciably greater than that of the
holes.

23.3 Semi-Conductor Junctions

On the basis of the theory outlined above, it
should not be surprising that a junction of semi-
conducting materials of opposite types can con-
stitute an excellent rectifier. When a battery,
poled as indicated in Figure 23-4, is connected
across such a junction, holes will be pulled out of
the p-type material toward the right and electrons
will be pulled out of the n-type material toward
the left. There can then be no current flow, except
that due to such relatively few minority carriers
as may be present (holes in the n-type material,
electrons in the p-type). When the applied voltage
is in this reverse direction, there are practically
no carriers at the junction which thus becomes an
effective insulator. It is of interest to note, how-
ever, that the immobile donor and acceptor atoms
remain, each of which has its own charge as illus-
trated in the drawing. Very high static electric
fields may therefore exist, resulting in potential
differences of considerable magnitude across the
junction.

If the polarity of the battery is reversed, on the
other hand, the forward direction of the applied
voltage causes both holes and electrons to move
toward the junction. Here they neutralize each
other. The net result is a free flow of electrons in
the external circuit which is limited only by the
numbers of carriers in the semi-conducting ma-
terials. Depending on this factor and the physical
size of the junction itself, n-p junction devices
may be designed to transmit quite high currents
in the forward direction and to give very high
rectification ratios.

On the basis of the theory outlined above, it
might be reasonable to assume that, in the ab-
sence of an external applied voltage, the electrons
of the n-type material and the holes of the p-type
material would gradually diffuse across the junc-
tion until the difference between the two semi-
conductor types was entirely destroyed. As a
matter of fact, this does not happen. When the
junction is formed, it may be assumed that such
diffusion of carriers across the junction starts.
The carriers of opposite sign cancel each other
out as they cross the junction, leaving on each side
a layer of immobile charged atoms, as indicated
in Figure 23-5. The electric fields set up by these
fixed charges prevent further diffusion of the
mobile carriers toward the junction. In other
words, a potential barrier is automatically estab-
lished across the junction between the two types
of semi-conducting material, allowing each to
maintain its own special characteristics. In the
typical p-n junction, the magnitude of this po-
tential barrier is in the order of tenths of a volt.
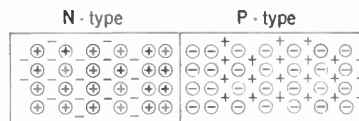
FIG. 23-5 POTENTIAL BARRIER AT P-N
JUNCTION

23.4 The Junction Transistor

The interesting behavior of the semi-conductor
junction can be taken advantage of to produce
another important electrical phenomenon known
as the transistor effect. The device pictured in
Figure 23-6 is a single crystal of germanium
which has been doped in such a way as to consist
of two blocks of n-type material separated by a
thin slab of p-type semi-conductor. It thus in-
cludes two p-n junctions. Separate electrical con-
nections may be made to the p-type section and
to each of the n-type sections, as shown in the
Figure. Now if we consider only the left junction
(i.e., with the circuit opened at E_c) it will be
noted that the battery E_c is poled in the forward
direction to permit the ready flow of a substantial
current, I_c. Electrons move freely from the left
n-material across the n-p junction into the p-
material, and holes from the latter into the former.
Considering only the right junction, on the other
hand, the battery E_c is poled in the reverse direc-

tion so that practically no current can flow across the p-n junction.

When both battery circuits are closed, an entirely new effect appears. Under the influence of E_c, electrons from the left n-section move freely into the p-section as before. Here a few of them may be neutralized by holes. But if the p-section is made thin enough, most of the electrons will diffuse right across the p-section and be caught up by the field in the right n-section produced by battery E_c. The result is a current I_c flowing in the circuit through E_c, and practically no current flowing in the external connection to the p-type section. I_c may be almost, but never quite, equal in value to I_e. The ratio of I_c to I_e, which is usually designated by the Greek letter, α, ranges between .95 and .99 in typical junction transistors. The energy that drives I_c comes from E_c and E_e but **the value of I_c is determined solely by I_e.** This is the transistor effect, and its similarity to voltage phenomena in the three-electrode electron

FIG. 23-7 COLLECTOR CHARACTERISTICS
OF N-P-N JUNCTION TRANSISTOR

The amplifying capability of the transistor depends upon the fact, emphasized above, that the value of the collector current, I_c, is determined by the value of I_e. As long as the positive biasing voltage applied to the collector is large enough to maintain a positive field in the collector, nearly all of the electrons coming from the emitter will move into the collector and the current I_c will maintain a constant value slightly less than I_e. This is illustrated by Figure 23-7 which gives the collector characteristic curves of a typical n-p-n junction transistor for several values of I_e. Note that the collector voltage V_c indicated here is the potential applied to the collector and is not necessarily the same as the voltage of the biasing battery, E_c. A large load resistance may be inserted in the collector circuit without affecting the current value, even though the IR drop across the resistance approaches quite closely to the value of E_c. This means that, with an appropriate value of E_c, a small varying signal voltage introduced in the emitter circuit, as indicated in Figure 23-8, will produce corresponding voltage variations across the load of much greater magnitude. Very substantial voltage and power gains are thus readily attainable.

FIG. 23-6 N-P-N JUNCTION TRANSISTOR

tube immediately suggests amplification possibilities.

In discussing the transistor further, it will be convenient to make use of certain nomenclature which has become standardized for historical reasons that will appear later. In the device of Figure 23-6, the n-type section at the left is called the **emitter**, the p-type section is the **base**, and the right n-section is the **collector**. The conventional schematic representation is shown at the right. Here, the emitter may be identified by the arrow head, which is pointed in the direction of conventional current flow—i.e., opposite to the direction of electron flow. It should also be noted that the arrows indicating the direction of current flow in the emitter and collector circuits of Figure 23-6 are also pointed in the conventional direction for current flow. This practice is followed throughout this text although in much of the earlier literature it has been the practice to point all such arrows toward the transistor elements regardless of the actual direction of current flow.

FIG. 23-8 COMMON BASE AMPLIFIER CIRCUIT

Junction transistors are also made with a p-n-p arrangement. Their characteristics are generally similar to those of the n-p-n transistor and the above discussion may be employed to explain their behavior by simply reversing the direction of current flow and the polarities of the biasing batteries, and substituting holes for electrons as the majority current carriers. It should be noted that the emitter current of the p-n-p transistor will also flow in the opposite direction and this will be indicated in the conventional diagram by pointing the emitter arrow head toward the base instead of away from it as in the n-p-n case.

Whichever type of transistor is being dealt with, it is imperative to keep in mind that the emitter must always be biased in the forward direction and the collector in the reverse direction. Application of biasing potentials not poled in accordance with this basic principle may destroy the transistor.

23.5 Metal to Semi-Conductor Junction

In the two preceding Articles, attention has been confined to the behavior of junctions between semi-conductor materials whose major conduction properties depend on the presence of different types of carriers (electrons or holes). Long before the invention of the transistor, however, it was well known that junctions between certain metals and certain semi-conductors, when properly formed, would behave as rectifiers. Thus, large area junctions between copper and copper-oxide and between selenium and cadmium, or a cadmium alloy, have been used for many years in the telephone plant and elsewhere as power rectifiers for converting alternating to direct voltages. In a smaller version, the copper-oxide rectifier is also extensively used as a varistor or diode in the modulator and demodulator circuits of carrier systems. Both copper-oxide and cadmium are p-type semi-conductors. There are numerous other metal to semi-conductor combinations that will produce rectification, including many in which the semi-conductor may be n-type.

The rectifying behavior of the metal to p-type semi-conductor junction is readily explained in terms of the relatively simple physical picture of the atomic and charge carrier structure within the semi-conductor that we have been using. Thus, if we apply a voltage in the reverse direction to such a junction, as shown in Figure 23-9, it may be considered that the mobile holes are drawn to
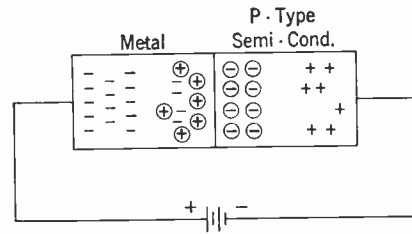
Fig. 23-9 Reverse Connection to Junction Between Metal and p-type Semi-Conductor

the right by the applied negative potential leaving at the junction an array of immobile atoms which are negatively charged. As in the n-p junction free electrons in the metal are drawn to the left, leaving positively charged metallic ions. The net result is a high potential barrier across the junction. The fields, quickly built up on either side of the junction, counteract the fields set up by the applied voltage and thus stop any further action.

As mentioned above, junctions between certain metals and certain n-type semi-conductors may also behave as rectifiers. This is notably true of the germanium diode which is usually built as a contact between tungsten and n-type germanium. The behavior of such a device does not lend itself to explanation in terms of the relatively simple physical picture of the atomic and current-carrier structure within the semi-conductor that we have employed up to this point. Thus in Figure 23-10, we may consider that the battery, connected in the reverse direction as shown, will draw the free electrons of the n-type semi-conductor toward the right and leave a layer of immobile positively charged ions at the junction. But what is to prevent the free electrons of the metallic conductor from crossing the junction and neutralizing the positive ions?

There is some potential barrier across the junction between many metals and semi-conductors and, for that matter, between various different kinds of metals. This is due to the fact that it requires the application of at least a certain energy to cause an electron to break through the surface and entirely out of the structure of which it normally forms a part. This energy is called the "work function" of the material and its value differs as between different metals or semi-conductors. When contact is made between conducting materials having different work functions, it is evident that conduction across the junction will be easier in one direction than in the other in proportion to the difference between the values of

[214]

work function of the two materials. This means, in other words, that there is an effective difference of potential across the junction. Such potentials, however, even in p-n junctions in semi-conductors, are only of the order of tenths of a volt. By themselves, they are insufficient to account for the failure to conduct when an external voltage of one or more volts is applied in the reverse direction.

To explain this behavior, the mathematical physicist resorts to the theory of electron energy bands. This theory derives largely from the Bohr theory of atomic structure which was in turn developed mainly from certain considerations of "quantum theory". The Bohr atomic theory postulates that the electrons surrounding the atomic nucleus can exist only in certain specific orbits, each of which is characterized by a specific energy level. Furthermore, only a specified number of electrons can be present in any one of the permitted energy levels. When originally suggested, this was a very bold hypothesis. However, later developments in the mathematical field of wave mechanics, which associates the energy of the moving electron with a specific wave frequency, gave it much support and made possible many quantitative predictions of atomic behavior that were subsequently verified by physical experimentation.

Employing similar mathematical reasoning, the energy band theory of conduction in solids postulates that free electrons in a crystalline solid can exist only at certain energy levels. Further, mathematical analysis indicates that the permitted levels occur in specific energy bands which, in some materials, are separated by "forbidden" energy gaps in which the existence of electrons is not possible. Electronic conduction in solids under normal conditions results from the movement of the carriers that may be present in the bands which include the higher energy levels, and all conduction is confined to carrier movement in these bands. In other words, only carriers that

have certain energy values can contribute to electrical conduction in the solid. Electronic conduction is generally confined to the band of highest energy levels, known as the **conduction band.** Conduction by means of positive holes may occur in the next lower energy band, known as the **valence band.**

Mathematical analysis under this theory enables the physicist to determine the "width" of the forbidden energy gap between the conduction and valence bands of materials in which he may be interested. In semi-conductors, the gap is substantial—about 0.7 electron volts for germanium and 1.1 electron volts for silicon. (An electron volt is the amount of energy required to raise an electron through a potential difference of 1 volt.) In good conductors, there is no forbidden gap, the conduction and valence bands in most cases overlapping one another. It follows that conduction electrons in a good metallic conductor are generally in lower energy levels than conduction electrons in the conduction band of a semi-conductor.

This difference of energy level explains the failure of the electrons in the metal of Figure 23-10 to move across the junction into the semi-conductor. Under normal conditions, they simply do not have enough energy to climb into vacant levels in the conduction band of the semi-conductor. Since there is no other place for them to go, no current can flow. It should be noted that the difference in energy levels is determined by the nature of the materials and is not affected by the value of the applied external voltage, up to a point of breakdown. If the external voltage is applied in the opposite (forward) direction, electrons from the conduction band of the semi-conductor can drop readily into the lower energy levels of the metal conduction band in accordance with the general law of nature under which any physical system tends always to adjust its total energy to the lowest possible value.

In the above, the energy band theory was introduced of necessity to explain the rectifying behavior of a metal to n-type semi-conductor junction. It is perhaps unnecessary to state that the theory applies equally well to the behavior of metal to p-type semi-conductor and p-n junctions, which were discussed in earlier Articles in more simplified terms. Indeed, the design of semi-conductor devices on other than a purely cut-and-try basis requires the use of this theory and its associated mathematics if the behavior of the device is
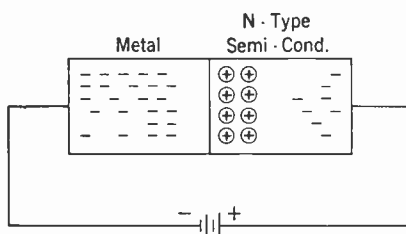
FIG. 23-10 REVERSE CONNECTION TO JUNCTION BETWEEN METAL AND N-TYPE SEMI-CONDUCTOR
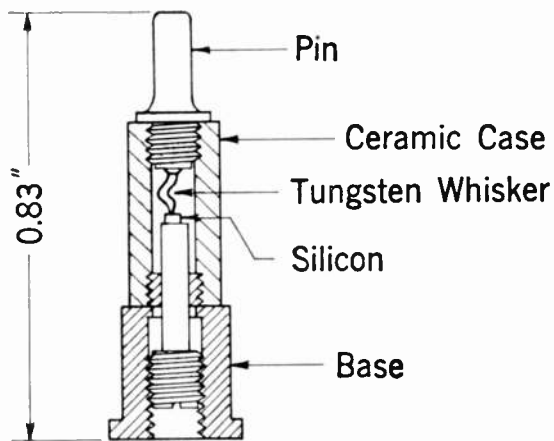
FIG. 23-11 POINT-CONTACT RECTIFIER

to be predictable in advance in quantitative terms.

23.6 Point-Contact Semi-Conductor Devices

The rectifying properties of metal to semi-conductor contacts were taken advantage of in the "crystal detector" which was a vital component of early radio receivers long before current theories of semi-conductor behavior had been developed. This device consisted of a metallic "cat-whisker" in contact with a small crystal of galena or iron pyrite. Its design was entirely empirical and practically no theory was available to explain its behavior. The modern descendant of this device is the point-contact rectifier or diode illustrated in Figure 23-11. This usually consists of a pointed tungsten or phosphor bronze wire in contact with a wafer of n-type germanium or silicon.

A completely satisfying explanation of the behavior of the point-contact rectifier, as distinguished from the large area metal to semi-conductor rectifiers described in the preceding Article, is still lacking. It is fairly obvious, however, that any fundamental difference in behavior must be accounted for by the great difference in the area of
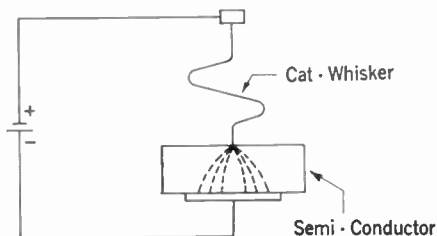
FIG. 23-12 ELECTRIC FIELD LINES IN POINT-CONTACT RECTIFIER

the contact junction itself. Thus, referring to Figure 23-12, it may be noted that if a forward voltage is applied across the junction an electric field is set up in the semi-conductor, whose lines of force are represented by the dashed lines. The intensity of this field, or the number of flux lines, is determined entirely by the value of the applied voltage. But all of the lines must terminate at the tiny point contact. The field intensity must therefore be very high in the immediate vicinity of the contact point. It is not unreasonable to assume that the extremely strong field in this area may then break many valence bonds of the semi-conductor and thus release a maximum number of electrons to flow into the cat-whisker. In any event, the fact is that the point-contact device gives much higher rectification ratios than can be obtained from large-area rectifiers. For obvious reasons, however, its current handling capacity is necessarily quite limited.
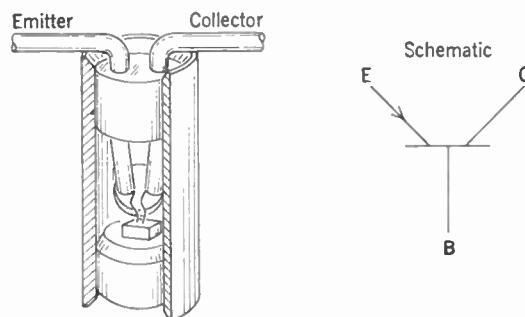
FIG. 23-13 POINT-CONTACT TRANSISTOR

More important than its relatively high rectification ratio is the fact that the point-contact device can handle a much broader band of frequencies. This can be explained by the difference in the effective capacitance of the different kinds of junctions. This capacitance, in the large area junctions, is relatively so large that the rectifying property of the junction is effectively lost at frequencies above a few tens of kilocycles in the case of the selenium rectifier, and above about two megacycles in the case of the small copper-oxide rectifiers that are used as modulators and demodulators in carrier systems. The much smaller capacitance of the point-contact rectifiers, on the other hand, permits them to handle frequencies as high as 1000 megacycles in the case of germanium and up to 10,000 megacycles in the case of silicon. This is the principal reason why such point-contact rectifiers are universally employed

as detectors in radar systems and in microwave radio systems.

It is an interesting fact that the original transistor of 1948 was also a point-contact device. As shown in Figure 23-13, this consisted of two cat-whiskers in contact with a tiny germanium block, the contact points being separated by from .001 to .002 inches. The emitter contact wire was usually of beryllium copper and the collector wire of phosphor bronze.

The behavior of this device might be expected to be very similar to that of the junction transistor discussed in Article 23.4. Holes injected into the germanium by a forward biased emitter would diffuse across the extremely small distance toward the collector point without too much danger of meeting an electron. Here they would be caught up in the field of the reverse biased collector. Thus, we might expect a collector current, I_c, which would be only slightly smaller than, but completely controlled by, the emitter current, I_e. However, the remarkable fact is that the collector current of the point-contact transistor takes values that are typically from 2.3 to 2.8 times as large as the value of the emitter current. Representative collector current-voltage curves for several values of I_e are given in Figure 23-14. These may be compared with like curves for the junction transistor shown in Figure 23-7.

No entirely satisfying explanation of this current-amplifying property of the point-contact transistor has been developed. Perhaps most acceptable is the hypothesis that the intense field in the vicinity of the emitter point creates so many holes that not all diffuse across to the collector. The excess holes, instead, are driven deeper into the body of the semi-conductor where they tend to build up a substantial positive field which can be neutralized only by an extra flow of electrons from the collector point. Most significant, however, is the fact that values of α appreciably greater than 1.0 are obtained only when the collector point contact has been subjected to a procedure known as "forming". This consists in driving relatively large but brief current pulses through the contact in the reverse direction. These produce momentary, very high temperatures at the junction point. It is thought that the brief temperature increase permits impurities from the phosphor bronze wire to diffuse into the semi-conductor, thus converting a tiny hemisphere directly under the point contact into p-type material. The net effect is to degrade the rectifying

behavior of the junction appreciably, which substantially increases the flow of reverse current in the collector circuit. It is considered that the additional electrons, thus flowing from the collector point produce a negative field in the direction of the emitter which encourages the flow of holes to the collector. These, in turn, require more electrons to flow from the collector for their discharge.

It is probable that further study of the behavior of point-contact devices will lead to a more com-
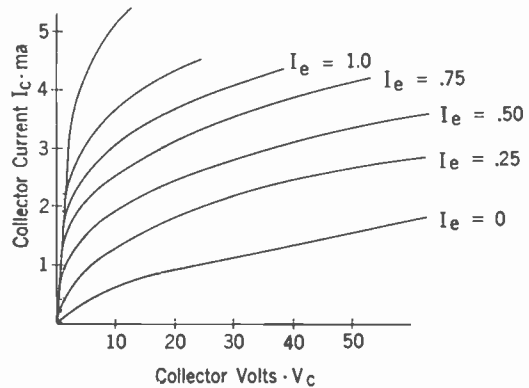
Fig. 23-14 Collector Current-Voltage Characteristics of Point-Contact Transistor

plete understanding of the phenomena involved in the behavior of the point-contact transistor. At the present time, this particular device is being largely superseded in most practical applications by junction transistors. This is a consequence of the development of technologies for manufacturing junction devices with good uniformity and precisely controllable characteristics. The point-contact transistor is still useful, however, in applications where the essential requirement is the effective, rapid opening and closing of circuits under the control of small signals from other circuits.

23.7 Semi-Conductor Photoelectric Devices

On the basis of the elementary theory of the behavior of semi-conductors, it is not surprising that such materials display photoelectric effects. Light of sufficient energy and appropriate wavelength will break valence bonds in the semi-conductor, thus liberating extra electrons and holes, which increase its conductivity. Accordingly, if a thin element of semi-conductor material such as selenium is arranged as in Figure 23-15, the current flowing in the external circuit as a result of the fixed biasing voltage will be increased by the

[217]
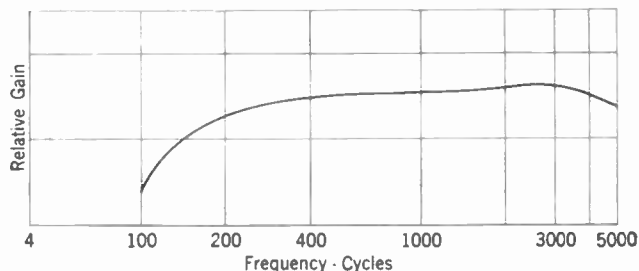
FIG. 24-3 GAIN-FREQUENCY CHARACTERISTIC OF AUDIO
AMPLIFIER

with a tetrode or pentode. However, since the principles involved are the same, it will be convenient in this and the following discussion to consider only triodes.

There is of course a limit to the total gain that can be obtained from a simple amplifier such as that of Figure 24-1, regardless of the type of tube used or the voltage step-up ratio of the input transformer. Where gains beyond this limit are required, a "cascade" amplifier may be used with the output of the first tube, or stage, working into the input of a second tube, and so on. Figure 24-2 shows a two-stage amplifier in which the output of the first tube is coupled to the input of the second tube through an inter-stage transformer. This transformer, as well as the input transformer, may have a substantial step-up ratio so that it too adds to the overall voltage amplification of the circuit.

The transformer coupled amplifier has limitations with respect to the frequency bandwidth that it can handle on a "flat" basis. This is due to the inductance of the transformer windings, and to their effective shunt capacitance. At frequencies below about 100 cycles per second, the inductive reactance of the primary winding of the inter-stage transformer is low enough in value so that the output resistance of the tube is not negligible in comparison with it. This results in a relative decrease in the voltage across the primary winding, and a consequent reduction in amplification. At frequencies above about 4000 cycles, the shunt capacitance becomes increasingly important. Since capacitive reactance is inversely proportional to frequency, the complex impedance of the transformer input becomes lower at the higher frequencies, with a consequent lower voltage across the primary winding. There is also a tendency to develop a resonance effect between the shunt capacitance and the inductance, which may produce a definite hump in the frequency-gain

curve near the higher frequency end. Well constructed amplifiers of this type, however, have a reasonably flat frequency response over a range from a little above 100 to approximately 5000 cycles. This is illustrated by the curve of Figure 24-3.

24.3 Resistance-Capacitance Coupled Amplifiers

Where flat frequency response over a greater range than four or five thousand cycles is required, resistance-capacitance coupling is commonly employed. An amplifier circuit of this type is shown schematically in Figure 24-4. Here, the a-c input is through the input capacitor C_i to a grid resistance R_g, the drop across which is applied to the grid of the first tube. Grid bias is provided by the drop across the cathode resistor R_k, through which the d-c component of the plate current flows. The a-c component of the plate current is by-passed by the capacitor C_k so that it
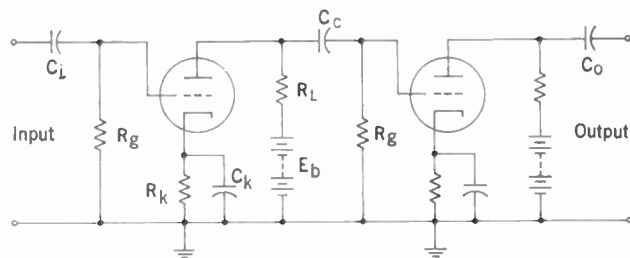
FIG. 24-4 RESISTANCE-CAPACITANCE COUPLED AMPLIFIER

has no effect on the grid. The alternating voltage drop across the resistor R_L is coupled to the input of the second tube by the capacitor C_c, which also prevents the plate battery voltage E_b from being impressed on the grid of the second tube.

An equivalent circuit for one stage of the R-C amplifier is shown in Figure 24-5. Here two shunting capacitors are indicated, which did not appear in Figure 24-4. C_{out} represents the inter-electrode capacitance of the first tube—chiefly the plate to cathode capacitance, C_{pk}—together
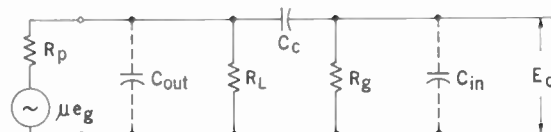
FIG. 24-5 EQUIVALENT CIRCUIT OF SINGLE STAGE OF R-C
COUPLED AMPLIFIER

[220]
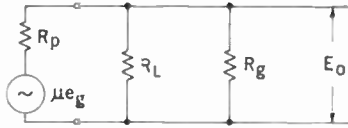
FIGURE 24-6

with such shunt capacitances as may be introduced by the circuit wiring. C_{in} represents a comparable capacitance C_{gk} at the input of the second tube. In a mid-frequency range—from about 100 to somewhat more than 10,000 cycles—the effect of the inter-electrode capacitances of the tubes is so small that the shunting capacitors C_{out} and C_{in} may be neglected. In this same frequency range, the coupling capacitor C_c may be considered as an a-c short-circuit because its reactance is negligible. The equivalent circuit then reduces to the parallel combination of resistors R_L and R_g across the tube output, as indicated in Figure 24-6. In the low frequency range, however, (below 100 cycles) the coupling capacitor C_c can no longer be ignored. Its reactance now becomes great enough that the voltage across R_L is divided between C_c and R_g with an increasing amount appearing across C_c, and a decreasing amount appearing across R_g. The input to the second tube is accordingly decreased, and the gain of the amplifier falls off. In the high frequency range (above 10,000 cycles), on the other hand, the effect of the coupling capacitor C_c again becomes negligible, but the net reactance of the shunting capacitors, C_{out} and C_{in}, then becomes small enough to cause the output voltage to fall off.

It will be clear that the bandwidth of reasonably flat frequency response of the R-C coupled amplifier depends mainly upon the values of C_c, R_L and R_g, and the values of the tube inter-electrode capacitances. Amplifiers of this type with a flat response over a range of 50 to some 15,000 cycles may be readily designed.

24.4 Video Amplifiers

There are many situations in modern communications practice where it is necessary for an amplifier to respond to a much broader band of frequencies than discussed above. Thus, the total frequency range of certain carrier systems extends over a band several million cycles in width. Even more demanding are certain transmission systems for short pulse signals and television signals. Here amplifiers are required to respond

without gain or delay distortion to frequencies ranging from as little as 20 cycles to more than 4 megacycles. Considerable modifications of the circuits described in the preceding Article are needed to meet these conditions.

It is necessary, in the first place, to employ tubes having maximum transconductance and minimum inter-electrode capacitance. In addition, the interstage coupling networks must be modified in such a way as to increase the gain in both the lower and upper frequency ranges. As might be expected, this can be accomplished only with some sacrifice of the mid-frequency gain. Coupling networks designed to "compensate" at both ends of the frequency scale may become quite elaborate, depending upon the degree of precision desired. The following indicates in a general way how the problem is approached. Considering first the lower frequencies, it was noted in connection with Figure 24-5 that the increased reactance of coupling capacitor C_c (since $X_c = 1/2\pi fC$) tends to decrease the voltage across R_g and hence the net gain. It is therefore desirable that C_c have the highest value practicable without unduly increasing its shunt capacitance to ground due to its physical size. The value of the cathode resistor by-pass capacitor, if used, should also be sufficiently large to make the $R_k C_k$ time constant long in comparison with the period of the lowest frequency. Beyond the limits of what can be accomplished by these means, the low frequency response can be further compensated by adding an equalizing network in series with the load resistance R_L, as shown in Figure 24-7. Here the values of R and C are so selected that the total impedance increases with decreasing frequency. The voltage applied to R_g and the amplifier gain is increased accordingly.

At the higher frequencies, the cause of decreasing gain is, as mentioned before, the increasing shunting effect of the inter-electrode capacitances of the tubes and of the wiring. As indicated in Figure 24-5 these capacitances, designated C_{out}
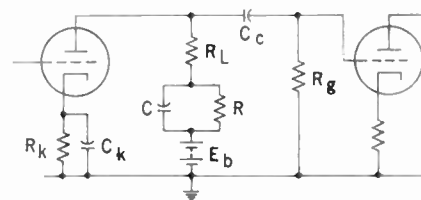
FIG. 24-7 LOW-FREQUENCY COMPENSATION

[221]
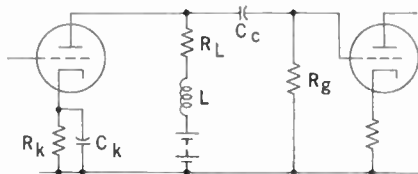
FIG. 24-8 HIGH-FREQUENCY COMPENSATION

and C_{in}, are in parallel with R_g. Their reactance decreases with increasing frequency, thus reducing the voltage across R_g and the gain. Compensation for this effect may be obtained by inserting an inductance in series with R_L, as shown in Figure 24-8. Since $X_L = 2\pi fL$, the inductor L causes the total load impedance to increase with increasing frequency. The voltage applied to R_g and the gain is increased proportionately. If this is insufficient to give complete equalization over the entire range, the gain at the higher frequencies may be further increased by bridging across L a capacitor of such value as to create a parallel resonant circuit that will further increase the total load impedance. A similar effect can be produced by adding a series inductor to resonate with C_c. Either or both of these latter refinements will of course increase the gain only over a limited part of the frequency range. But if the video amplifier includes several stages, as most of them necessarily do, each of the several interstage networks can be designed to peak the gain at a different point in the frequency range so that a reasonably uniform overall gain response results.

It is of interest to note that low-frequency and high-frequency compensation can be combined in a single interstage network if the low-frequency compensator is connected on the low-potential side of the high-frequency compensator, as shown in Figure 24-9. Neither compensator has any appreciable effect on the operation of the other because at very low frequencies, the impedance of L is so low as to make it practically a short circuit while at high frequencies, the very low impedance of C effectively shorts out the low-frequency network.
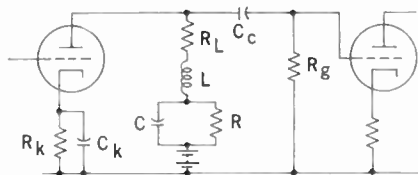
FIG. 24-9 COMBINED HIGH & LOW FREQUENCY COMPENSATION

24.5 Radio-Frequency Amplifiers

Voltage amplifiers for most of the applications in radio circuits, such as radio receivers and the low-power stages of radio transmitters, do not have to meet as severe requirements with respect to frequency response as do audio or video amplifiers. This is because the typical radio circuit is designed in theory to handle only the single frequency to which it is tuned. Actually, of course, the tuning is not so sharp that it does not permit the passage of a band of frequencies extending far enough on both sides of the tuned frequency to carry the complete communication signal. Coupling between the stages of radio amplifiers is commonly accomplished by means of simple air-core transformers. As indicated in Figure 24-10, one or both windings of the coupling
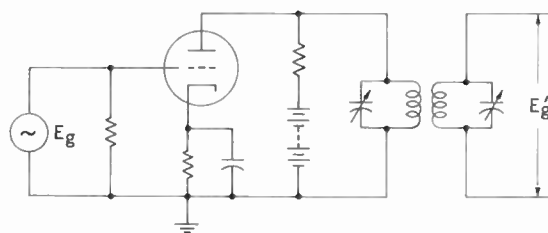
FIG. 24-10 RADIO AMPLIFIER CIRCUIT

transformer are tuned with a paralleling capacitor to the signal frequency. As discussed in Article 14.6, where both primary and secondary are so tuned, a good band of frequency response with sharp cutoff at each end is readily obtained. The transformer itself need have little or no voltage gain because tubes with high amplification factors are used.

24.6 Negative-Feedback Circuits

For a great majority of amplifier applications in telephone work, it is important not only that the output signal be a faithful reproduction of the input signal, but also that maximum stability of amplifier operation be secured. Both of these objectives can be met to a very large degree by feeding back some of the amplifier output to the input circuit in an inverse phase relationship. An amplifier so connected is called a **negative feedback** or degenerative amplifier. Its principle may be understood by referring to Figure 24-11. In this Figure, (a) indicates an amplifier without feedback, having an overall voltage amplification of
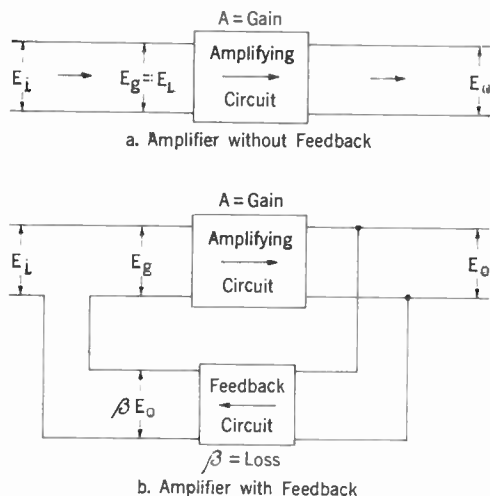
FIG. 24-11 PRINCIPLE OF NEGATIVE FEEDBACK

gain of A. In Figure 24-11 (b), a part of the output voltage is returned to the input, 180° out of phase with the input voltage, through a feedback circuit having a loss β. Without feedback (Figure 24-11 (a)) we have—

$$E_0 = AE_i = AE_g \qquad (24:1)$$

In Figure 24-11 (b), on the other hand, the actual input voltage E_g of the amplifier unit is no longer equal to E_i, but to the sum of this voltage and the feedback voltage βE_0. That is—

$$E_g = E_i + \beta E_0 \qquad (24:2)$$

The output voltage therefore is—

$$E_0 = AE_g = A(E_i + \beta E_0)$$

Solution of this equation for the overall voltage gain, E_0/E_i, gives—

$$\frac{E_0}{E_i} = \frac{A}{1 - \beta A} \qquad (24:3)$$

Since the feedback is inverse (negative), the value of βA is negative and the denominator of equation (24:3) is greater than unity. Negative feedback accordingly always reduces the net gain, but the reduction can be compensated by the use of an amplifier having as high gain A as may be required to obtain the desired overall gain. When the product βA is much larger than unity, as is the case in most practical circuits, the overall amplification becomes effectively—

$$\text{Net Amplification} = -\frac{1}{\beta} \qquad (24:4)$$

In other words, the effective gain of the circuit

depends entirely upon the characteristics of the feedback circuit. This may perhaps be better understood by considering a numerical example. In the circuit of Figure 24-12, the gain A of the amplifier unit is 80 db (voltage ratio of input to output of 1 to 10,000) and the loss in the feedback circuit is 60 db (voltage ratio of 1000 to 1). From equations (24:2) and (24:3)—

$$E_g = E_i \left(\frac{1}{1 - \beta A} \right) \qquad (24:5)$$

If the applied input voltage E_i is 1 millivolt, the actual input voltage to the amplifying unit is therefore—

$$E_g = \frac{1}{1 - \dfrac{10,000}{-1,000}} = \frac{1}{1 - (-10)} = \frac{1}{11}$$

$$= .09091 \text{ millivolt}$$

The output voltage is—

$$AE_g = E_0 = 10,000 \times .09091 = 909.1 \text{ millivolts}$$

This output of 909.1 millivolts is also impressed
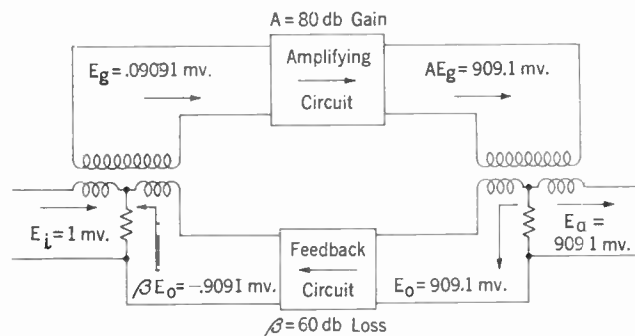
FIG. 24-12 EXAMPLE OF NEGATIVE FEEDBACK CIRCUIT

on the feedback circuit which allows 1/1000 of it to be fed back to the amplifier input. In passing through the feedback circuit its phase is shifted until it is out of phase with the applied input of 1 millivolt, which gives it a minus sign. We then have $-.9091$ millivolt combining with the initial 1 millivolt to give the actual input voltage to the amplifier, which, therefore, is—

$$1.000 - .9091 = .0909 \text{ millivolt}$$

This checks the value of E_g obtained above, which means that the amplifier is stable and as long as the applied input of 1 millivolt is maintained, there will be 909.1 millivolts in the output. The

[223]

overall gain of the amplifier under these conditions is—

$$20 \log_{10} \frac{\text{Output voltage}}{\text{Input voltage}} = 20 \log_{10} \frac{909.1}{1}$$

$$= 20 \times 2.9586 = 59.17 \text{ db}$$

It will be noted that for all practical purposes this gain is the same as the loss β of the feedback circuit.

If we had used an amplifier unit with a higher gain—say 100 db (voltage ratio of input to output of 1 to 100,000)—and the same loss in the feedback circuit, we might expect the output voltage to be much higher, but such is not the case. Using equation (24:5) again we find the actual input voltage now is—

$$E_g = \frac{1}{1 - \dfrac{100,000}{-1,000}} = \frac{1}{101} = .009901 \text{ millivolt}$$

The output voltage, E_0, accordingly is—

$$100,000 \times .009901 = 990.1 \text{ millivolts}$$

and overall gain of the amplifier is—

$$20 \log_{10} \frac{990.1}{1} = 20 \times 2.9957 = 59.91 \text{ db}$$

which is again practically equal to the loss in the feedback circuit. This means that **even if the gain A of the amplifier unit changes due to variations in the battery supply, changing tube characteristics, etc., the overall gain remains the same for all practical purposes.**

Another important feature of the negative feedback amplifier is its ability automatically to reduce to a negligible magnitude any noise or harmonic distortion developed within the amplifier itself. This is true because a part of this noise and distortion appearing in the output is fed back to the input through the feedback circuit where it re-enters the amplifier in such a phase relation that when it is amplified and again appears in the output, it is out of phase with the original noise
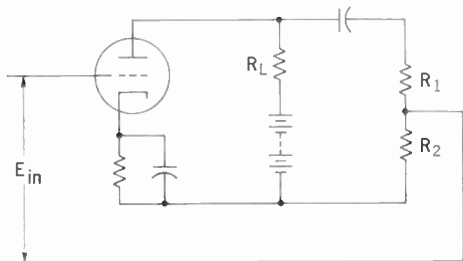
FIG. 24-13 VOLTAGE FEEDBACK CIRCUIT
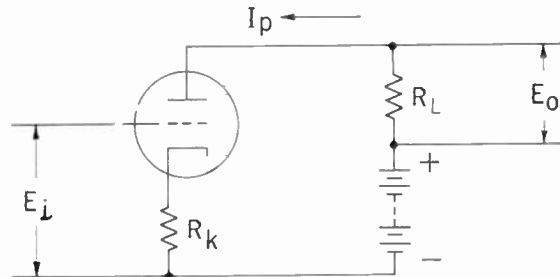
FIG. 24-14 CURRENT FEEDBACK CIRCUIT

and distortion, thereby reducing its effect.

Feedback circuits may be designed with either voltage feedback or current feedback. Figure 24-13 illustrates a simple voltage feedback arrangement. Here the total resistance $(R_1 + R_2)$ of the voltage divider is made large enough so that its shunting effect on the load resistance R_L is practically negligible. The magnitude of the feedback factor β is—

$$\beta = \frac{R_1}{R_1 + R_2}$$

and the overall gain when βA is large compared to unity is the reciprocal of this—

$$\text{Net gain} = \frac{R_1 + R_2}{R_1} \qquad (24:6)$$

The voltage fed back is 180° out of phase with the alternating input voltage because it represents a voltage drop in the plate circuit, and the plate current is in phase with the input voltage.

The simplest type of current feedback arrangement may be obtained by the use of an unbypassed cathode resistor, as indicated in Figure 24-14. Here the alternating plate current I_p must flow through R_k as well as through the load R_L. This causes a voltage drop across the cathode resistor equal to $I_p R_k$, and the net input voltage applied to the grid is then—

$$E_g = E_i - I_p R_k$$

This may be rewritten in terms of E_0 as follows—

$$E_g = E_i + \frac{E_0 R_k}{R_L}$$

since E_0 is equal to $-I_p R_L$, the drop across the load resistance. Comparison of this equation with (24:2) shows that $\beta = R_k/R_L$; and when βA is large compared to unity—

$$\text{Net gain} = \frac{R_L}{R_k}$$

[224]

24.7 The Cathode-Follower

An interesting example of maximum application of negative feedback is displayed in the so-called **cathode-follower** circuit, shown in Figure 24-15. Here there is one hundred percent current feedback through the cathode resistor, R_k—in other words, β is equal to unity. The output is taken across the cathode resistor so that the input and output voltages are necessarily in phase and the a-c cathode to ground voltage varies in the same direction or "follows" the applied grid to ground voltage. The net voltage gain of the circuit is always less than unity because of the hundred percent negative feedback. It may be expressed as—

$$\text{Net gain} = \frac{\mu R_k}{R_p + R_k (1 + \mu)}$$

where μ is the amplification factor of the tube and R_p is its plate resistance.

As an amplifier, the cathode-follower circuit would appear useless since its voltage gain is less than one. However, it is still capable of delivering power to a load without requiring appreciable input power, and with extremely faithful reproduction of the variations in the input voltage. The circuit is very stable and virtually independent of any variation in the tube characteristics. These factors make it useful as a stabilizing coupling circuit between an amplifier and a load. More important is the fact that while the input impedance of the circuit is high, its output impedance is very low for an amplifier. This output impedance consists of R_k in parallel with an effective plate resistance equal to $R_p/(1 + \mu)$. Since the value of $R_p/(1 + \mu)$ is less than 1000 ohms for most tubes, the value of the net output impedance must be still less. As an impedance-matching device accordingly, the cathode-follower circuit is useful for such purposes as coupling the relatively high impedance output of a video amplifier to the low impedance of a coaxial line.
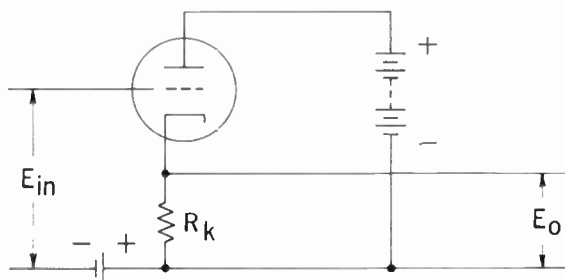
24.8 Power Amplifiers

The classification of amplifier circuits as between voltage amplifiers and power amplifiers is not very definite. Actually, the term "power amplifier" is somewhat misleading because it is perfectly possible for an amplifier to deliver a substantial power output without appreciable power input. The power classification is applied generally in practice to situations where the delivery of a desired amount of power is the controlling criterion, and voltage gain, if any, is of secondary importance. Thus, an amplifier used to drive a load such as a loudspeaker or other device requiring considerable power for its operation, is ordinarily classed as a power amplifier. So is the amplifier
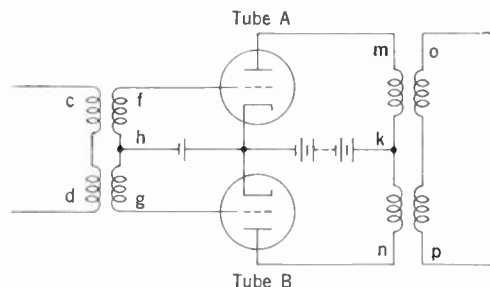
Fig. 24-16 Push-Pull Amplifier

that must supply many kilowatts of power to drive the antenna of a radio transmitter.

The power that a vacuum tube amplifier can develop of course depends generally on the maximum value of current that may flow in its plate circuit. In most audio amplifier applications, this value is limited by the fact that the tube must operate on the straight line portion of its characteristic curve in order that its output be a faithful reproduction of the input signal. The maximum power that can be delivered, therefore, depends upon the size and characteristics of the tube or tubes used. Where the power requirement is a matter of a very few watts, as for driving the speaker of an ordinary radio receiver, a single triode or pentode may be used. Such a tube is not different in appearance from the tubes used in voltage amplifying circuits, although its design characteristics will generally be such that it will have a lesser amplification factor and a larger plate current.

Fig. 24-15 Cathode Follower Circuit

[225]

Where more power than a single tube can deliver is required, together with maximum fidelity of signal reproduction, two tubes may be employed in a "push-pull" circuit, as shown in Figure 24-16. In this circuit, the two tubes, *A* and *B*, have identical characteristics. An alternating voltage applied at the input, *cd*, impresses voltages of equal magnitude but opposite polarity upon the control grids of tubes *A* and *B*. As the control grid of one tube becomes less negative (more positive), its plate current increases; at the same time, the control grid of the other tube becomes equally more negative, which decreases its plate current; and vice versa. Since the plate battery is connected to the midpoint, *k*, of the primary winding, *mn*, of the output transformer, the plate currents flow in opposite directions in each

A - A-C Components in Winding m · k, from Tube A

B - A-C Components in Winding n · k, from Tube B

C - A-C Components in Output Winding, o · p, from Tubes A & B

FIG. 24-17 PRINCIPLE OF PUSH-PULL AMPLIFIER

half of the primary winding. When the two plate currents are equal, therefore, there is no current in the secondary winding, *op*. On the other hand, a decreasing plate current in one half of the primary winding, and an increasing plate current in the other half, induce equal currents in the same direction in the secondary winding, *op*. The total output is thus obviously equal to the sum of the outputs of the two tubes.

As a matter of fact, the push-pull circuit will provide a power output appreciably greater than twice the output of an amplifier employing only one tube. This is due to the fact that the tubes of the push-pull amplifier may be given more control grid bias than a single tube without causing distortion in the output. The characteristic curve of a triode is of such shape that its output, when working over a portion of the curve including some curvature, consists principally of the fundamental or desired frequency, and its second harmonic (double the fundamental frequency). Such outputs are illustrated in Figure 24-17 (A) and (B) where it will be noted that the net output of each tube, represented by the heavy lines, is considerably distorted. It may also be noted, however, that the second harmonics in the outputs of both tubes become positive and negative at the same time. This means that the components of the current represented by these harmonics are always flowing in opposite directions in the halves of the primary winding, *mkn*, and accordingly produce no effect in the secondary winding, *op*. In other words, the second harmonics cancel each other. The net result is indicated in Figure 24-17 (C), where the two output currents are shown to add to produce a sine wave, which is a faithful reproduction of the input signal. Because the push-pull amplifier can thus be operated over a greater range of its tubes' characteristic curves, its output may actually exceed by more than three times the equally distortionless output that could be obtained from a single tube amplifier.

All of the amplifiers that have been considered thus far have operated on a high fidelity basis— i.e., so that the output signal presents a faithful reproduction of the input signal. In radio parlance, such operation is designated "Class A", a term which indicates, in general, that the amplifier tubes are operating only on a straight line portion of their characteristic curves. All Class A amplifiers have comparatively low "plate efficiency"—that is, the ratio of their useful output power to the total power supplied to the plate cir-

Class B

A

Class C

B
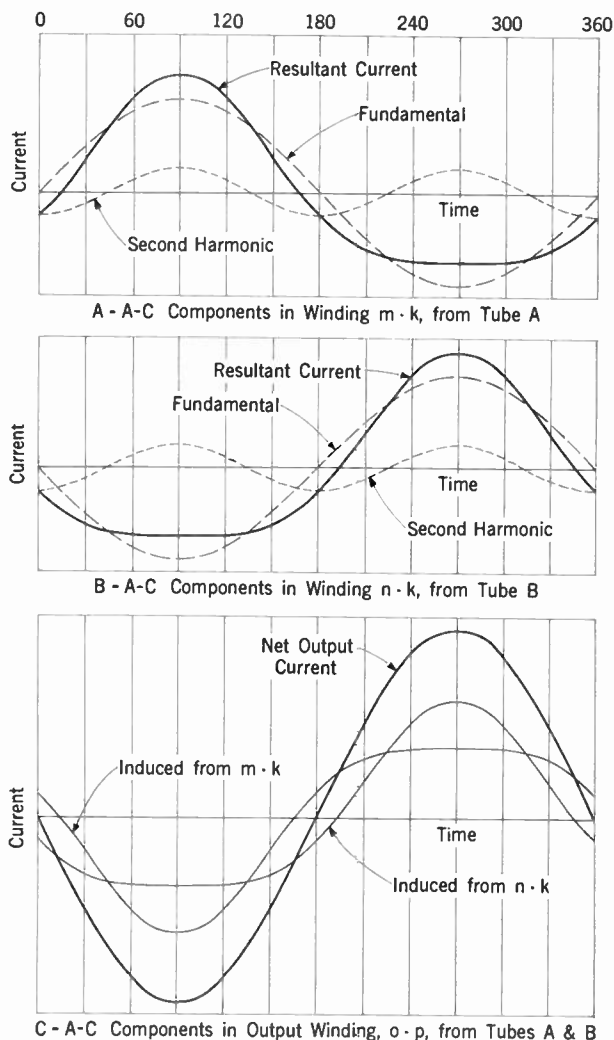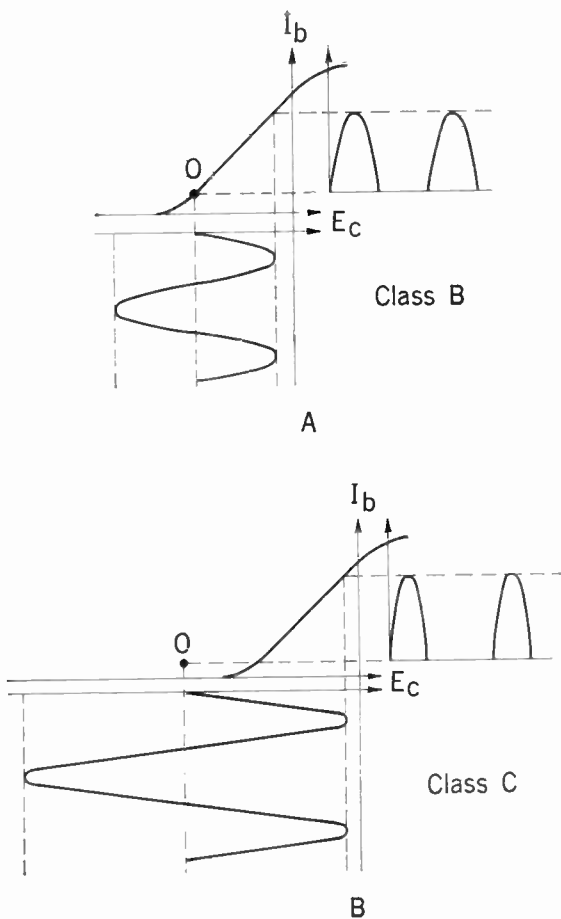
FIG. 24-18 OPERATING CHARACTERISTICS OF CLASS B
AND C AMPLIFIERS

cuit by the B battery cannot be greater than 50%
in theory, and is usually not much higher than
25% in practice. In audio amplifiers, where the
output power is in any event not very great, this
is not too important. In high-powered radio trans-
mitters, on the other hand, where output power is
measured in kilowatts, better efficiency becomes
economically significant, and leads to the use
when possible of "Class B" and "Class C" opera-
tion. The grids of Class B amplifiers are biased
to the cutoff point so that plate current flows only
during one-half of the cycles of an applied alter-
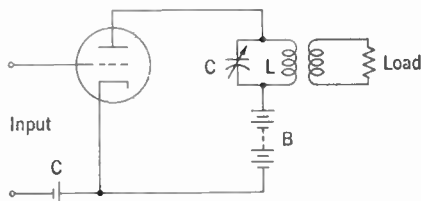
FIG. 24-19 R-F POWER AMPLIFIER

nating voltage. In Class C operation, the grid is
biased well beyond the cutoff point so that plate
current flows during less than one-half of each
cycle of applied grid voltage. The effective re-
sults are indicated in Figure 24-18 (A) and (B)
respectively. It is evident that in both cases, the
output wave form presents a highly distorted ver-
sion of the wave form of the input signal. How-
ever, plate current flows and draws power from
the B supply only part of the time as contrasted
with the continuous power drain in Class A oper-
ation. The plate efficiency of these types of ampli-
fiers is therefore higher, having a theoretical pos-
sible maximum of 78% in the case of the Class B
operation, and as much as 85% in Class C opera-
tion.

Radio power amplifiers ordinarily work into a
load impedance which includes a parallel-tuned
circuit, as indicated in Figure 24-19. This L-C
tank circuit, when tuned to the operating fre-
quency, acts as a filter to suppress the many har-
monics of the fundamental frequency that must
obviously be present in the plate circuit of a tube
operating Class B or C. The current flowing in
the load itself consequently represents only the
fundamental input frequency. Its amplitude,
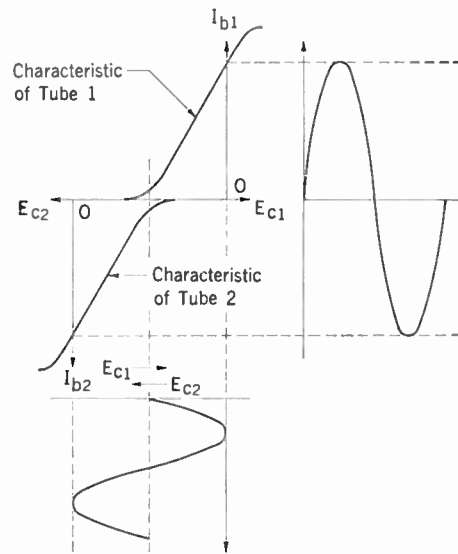
FIG. 24-20 GRAPHICAL REPRESENTATION OF CLASS B
PUSH-PULL AMPLIFIER OPERATION

however, cannot be expected to be linearly re-
lated to the input voltage. In Class B operation,
distortion may be held to reasonable proportions
by operating two tubes in a push-pull arrange-

[227]
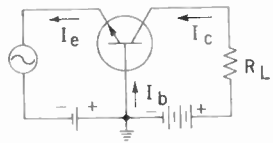
FIG. 24-21A GROUNDED BASE TRANSISTOR AMPLIFIER
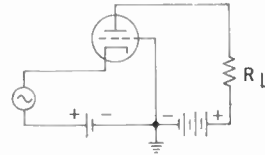
FIG. 24-21B GROUNDED GRID VACUUM TUBE AMPLIFIER

ment. In this case, one tube will provide an approximately true reproduction of the positive half of the input voltage wave, while the other tube furnishes a like reproduction of the negative half of the input wave. The net effect is illustrated graphically in Figure 24-20.

24.9 Transistor Amplifiers

The performance of all of the amplifier circuits discussed in the preceding Articles of this Chapter can be closely simulated by circuits employing junction transistors instead of vacuum tubes as their active elements. Such circuits are generally similar in design to the vacuum tube circuits, although certain unique characteristics of the transistor, such as its sensitivity to temperature changes, may necessitate some rather fundamental differences in the detailed circuitry.

In Chapter 23, the amplifying ability of the junction transistor was explained by means of an n-p-n transistor connected as shown in Figure 23-8 with its base common to the emitter and collector circuits. This is known as a "common base" or "grounded base" connection. The input signal is applied to the emitter and the output signal is taken from the collector. Input and output signals are in phase. Current amplification can only approach 1.0 but substantial voltage and power gains are possible because the input impedance is low (in the order of 50 ohms) and the output impedance is very high (in the order of .5 megohms). Figure 24-21a is a conventional diagram of a grounded base amplifier circuit employing an n-p-n transistor; Figure 24-21b illustrates its

counterpart, a grounded grid vacuum tube amplifier circuit. Both circuits can give high voltage amplification, both amplify without changing the phase of the signal voltage, and both have low input and high output impedance.

The grounded grid vacuum tube circuit has limited application as an amplifier in telephone practice. This is also true of the grounded base transistor circuit although it may have many radio frequency applications. Generally preferred at the lower frequencies is the grounded or common emitter circuit which is analogous to the familiar grounded cathode vacuum tube circuit. Figures 24-22a and b show the essentials of these two circuits. It will be noted that the input signal is applied to the base of the transistor amplifier and to the grid of the vacuum tube. In both circuits the phase of the output signal is inverted with respect to the input. The grounded emitter connection can provide current amplification as high as 50 or more. This would appear to contradict the fact that the α value of the junction transistor cannot exceed one. However, α is the ratio of collector current to emitter current whereas the current gain in this connection is represented by the ratio of collector current to the much smaller base current. This ratio, which is sometimes designated β, may be evaluated in terms of α by noting that

$$I_c = \alpha I_e \qquad (24:7)$$

and that

$$I_b = I_e - I_c = I_e - \alpha I_e = I_e (1 - \alpha) \quad (24:8)$$

The current amplification ratio then is—
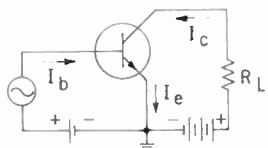
FIG. 24-22A GROUNDED EMITTER TRANSISTOR
AMPLIFIER
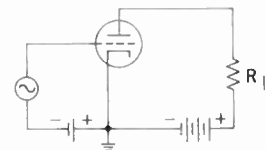
FIG. 24-22B GROUNDED CATHODE VACUUM TUBE AMPLIFIER
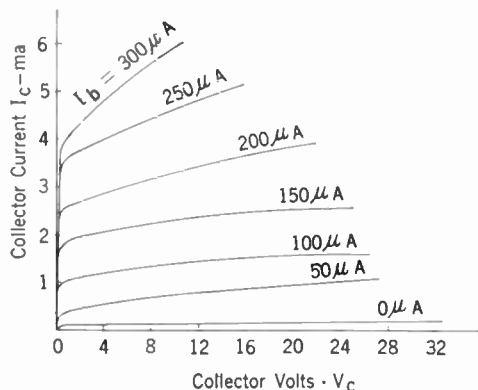
FIG. 24-23 COLLECTOR CHARACTERISTICS-GROUNDED EMITTER CONNECTION

$$\frac{I_c}{I_b} = \beta = \frac{\alpha I_e}{I_e (1 - \alpha)} = \frac{\alpha}{1 - \alpha} \qquad (24:9)$$

From this, it is evident that the current gain is still determined by α and can be very large when the value of α is close to 1.0.

Representative collector voltage-current curves for a grounded emitter amplifier at various values of base current are given in Figure 24-23. Comparing these curves with the characteristics for grounded base operation, as given by Figure 23-7, it may be noted that the grounded emitter characteristics have appreciably greater slope. The output impedance is accordingly lower—in the order of 30,000 to 50,000 ohms. The input impedance, on the other hand, is considerably higher than that of the grounded base connection, ranging typically from 700 to 1000 ohms. Despite the smaller ratio between input and output impedance, the grounded emitter circuit is generally capable of giving higher power gain than the grounded base circuit because of its substantial current amplification. Typical power gain is in the order of 5000 or more (35-40db).

A third possible transistor amplifier connection is illustrated in Figure 24-24a. This is known as a common or grounded collector circuit. It is analogous to the grounded plate vacuum tube
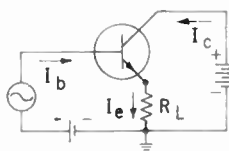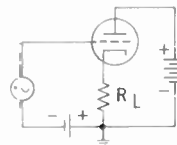
circuit shown in Figure 24-24b, which will be recognized as the cathode-follower discussed in Article 24.7. Characteristics of the grounded collector transistor circuit are very similar to those of its vacuum tube counterpart. It has high input impedance and low output impedance but is capable of giving some power gain although its voltage gain is always less than 1. Like the cathode-follower, it is used principally for impedance matching purposes.

The grounded emitter connection is used almost universally in all audio applications of the transistor, while the grounded base as well as the grounded emitter configuration are used in *RF* and *IF* amplifier applications. It is necessary to note that the grounded emitter connection is inherently less stable than the grounded base connection. This is due to the effect of the small reverse current which is always present in the collector-base circuit whether or not there is current in the emitter circuit. The value of this current, usually designated I_{co}, is normally only a few micro-amperes but it is extremely sensitive to any temperature change at the collector-base junction.

In any transistor circuit connection, I_{co} is a part of the total collector current I_c. With the grounded base connection, I_c is equal to αI_e plus I_{co}. But since α is at least .95 and I_c is at least several milliamperes, I_{co} can be neglected. Even a large increase in the temperature of the collector junction could not increase the value of I_{co} enough to make it of any consequence as compared with αI_e. With the grounded emitter connection, on the other hand, the collector current value is normally determined by the base current I_b which is equal to a small fraction of I_c (less than 5%) plus I_{co}. When α is large, I_{co} forms a significant fraction of the total base current. A relatively small increase in its value due to temperature change may therefore cause a considerable increase in the collector current when the increase is multiplied by the amplification factor β of Equation (24:9). This may be sufficient to further increase the

FIG. 24-24A GROUNDED COLLECTOR TRANSISTOR AMPLIFIER

FIG. 24-24B GROUNDED PLATE VACUUM TUBE AMPLIFIER

[229]
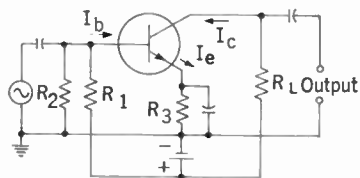
FIG. 24-25 STABILIZED GROUNDED EMITTER CIRCUIT

temperature at the junction, thus setting up a cumulative process that will result in serious distortion of the characteristics of the transistor, if not in its complete destruction.

To counteract the tendency toward instability in the grounded emitter amplifier circuit, it is the usual practice to insert a resistor in series with the emitter. This is indicated as R_3 in Figure 24-25. Collector current flowing in this resistance produces a voltage drop that makes the emitter positive with respect to ground. The net base-emitter voltage is then equal to the difference between the positive potential applied to the base by the battery through resistors R_1 and R_2 and the drop across R_3. Then, an increase in I_{co} due to temperature change will increase the drop across R_3 which will reduce the base-emitter voltage accordingly. This decreases the value of I_b and I_c, thus counteracting the increase in I_c. Typical values for the stabilizing resistor R_3 are in the order of 5000 ohms while corresponding values for R_1 and R_2 would be around 10,000 ohms. The stabilizing resistor may be by-passed with a capacitor, as shown in the drawing, to avoid degenerative feedback. Figure 24-25 also shows how a single potential source can be used to bias both emitter and collector—negative at the emitter and positive at the collector, as is necessary for the n-p-n transistor.

In general, the circuitry of multi-stage or cascaded transistor amplifiers is closely similar to that of comparable devices employing vacuum tubes. Both transformers and resistance capacitance networks are used for interstage coupling and negative feedback is commonly employed to insure stable overall operation. The fact that junction transistors are of two different kinds, however, permits certain amplifier designs that have no counterpart in vacuum tube circuits. One such circuit is illustrated in Figure 24-26 where it will be noted that an n-p-n transistor is used in the first stage and a p-n-p in the second. This permits direct coupling without a series capacitor because the first-stage load resistor R_1 is assigned a value such that the collector current flowing through it produces a voltage drop that makes the end away from the battery just negative enough to keep the base of the second-stage transistor negative with respect to the emitter. The emitter circuit of the p-n-p transistor is thus given a forward bias enabling it to further amplify the already once amplified signal voltage developed across R_1 by the first transistor.
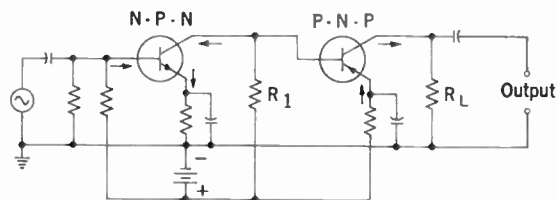
FIG. 24-26 DIRECT-COUPLED AMPLIFIER WITH COMPLEMENTARY TRANSISTORS

All of the circuits discussed above are essentially voltage amplifying devices employing small transistors having maximum power outputs generally measured in milliwatts. It needs to be noted that transistors suitable for use in power amplifiers are also available. These can deliver output power ranging up to 10 watts or more and are commonly used in push-pull connections similar to the power amplifiers of vacuum tube practice. The elements of power transistors are naturally larger and they usually operate with higher biasing voltages. They are equally sensitive to temperature and require the use of special measures to dissipate the heat generated at the collector by the relatively high output currents. Metal cooling fins, water jacketing, and forced air cooling are variously employed for this purpose.

[230]

CHAPTER 25

AMPLIFIER APPLICATIONS IN TELEPHONE PRACTICE

25.1 Telephone Repeaters

Modern telephone practice requires the use of nearly all types of amplifiers to meet its various needs. An important application is the relatively simple voice-frequency amplifier commonly known as the telephone repeater. This device is inserted at regularly spaced intervals in long voice-frequency telephone circuits to offset line attenuation.

Since telephone conversation moves in both directions, each such telephone repeater must include two oppositely poled amplifiers. In a two-wire telephone circuit, accordingly, the telephone repeater must be "double-tracked" by means of hybrid coils or comparable devices, as was dis-

cussed in Article 14.7. Figure 25-1 is a schematic of a telephone repeater known as the 22-type. As will be observed, the amplifier units in this arrangement are triodes connected with transformers in both input and output circuits. The drawing also shows the connections of the hybrid coil output transformers with their balancing networks; and the potentiometers in the input circuits for controlling the amplifier gains. Equalizing networks are inserted at the mid-points of the low-impedance sides of the input transformers; and low-pass filters are included in the output circuits to prevent the passage of high frequencies not essential for voice transmission. The maximum
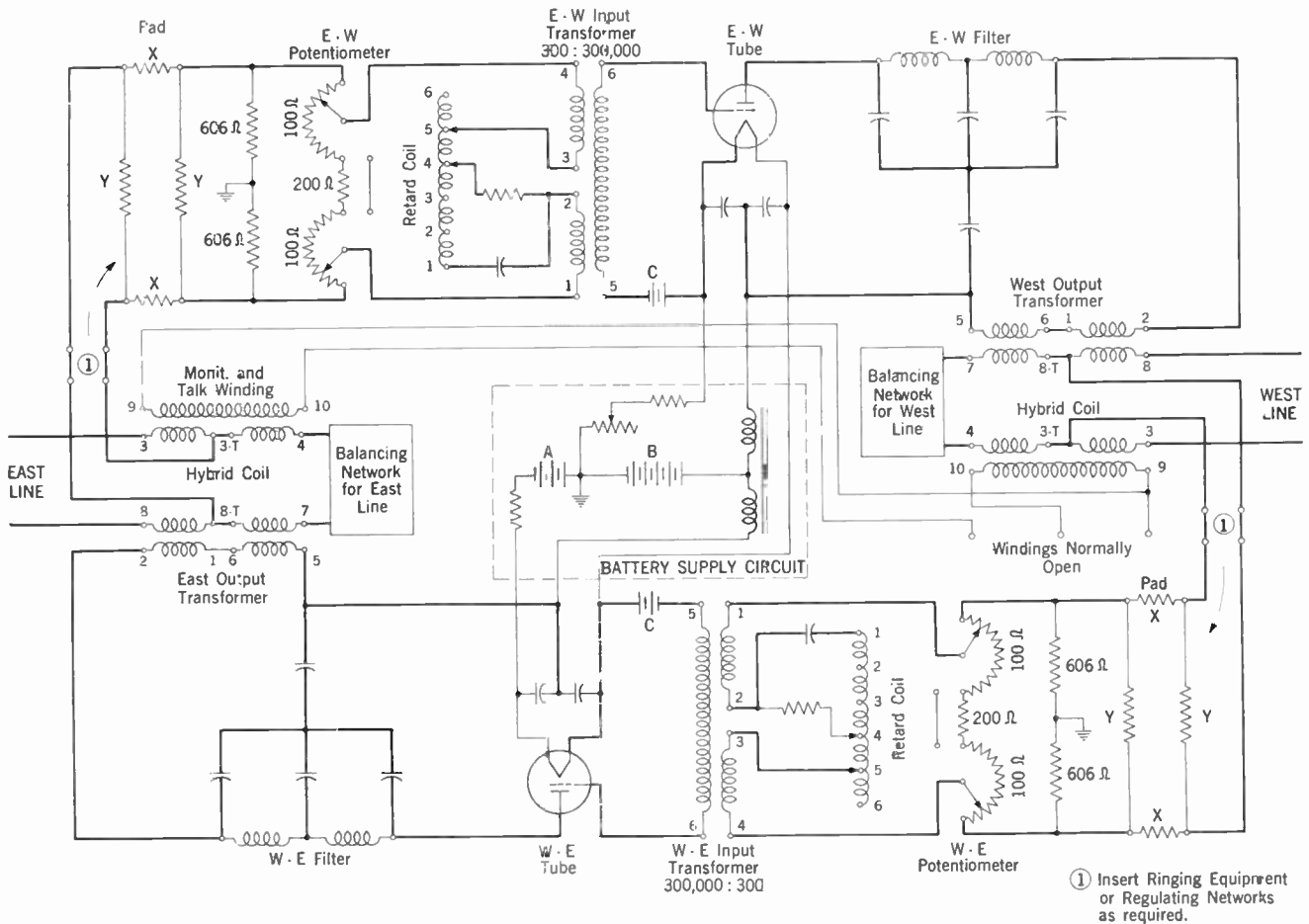
FIG. 25-1 22-TYPE TELEPHONE REPEATER CIRCUIT

[231]
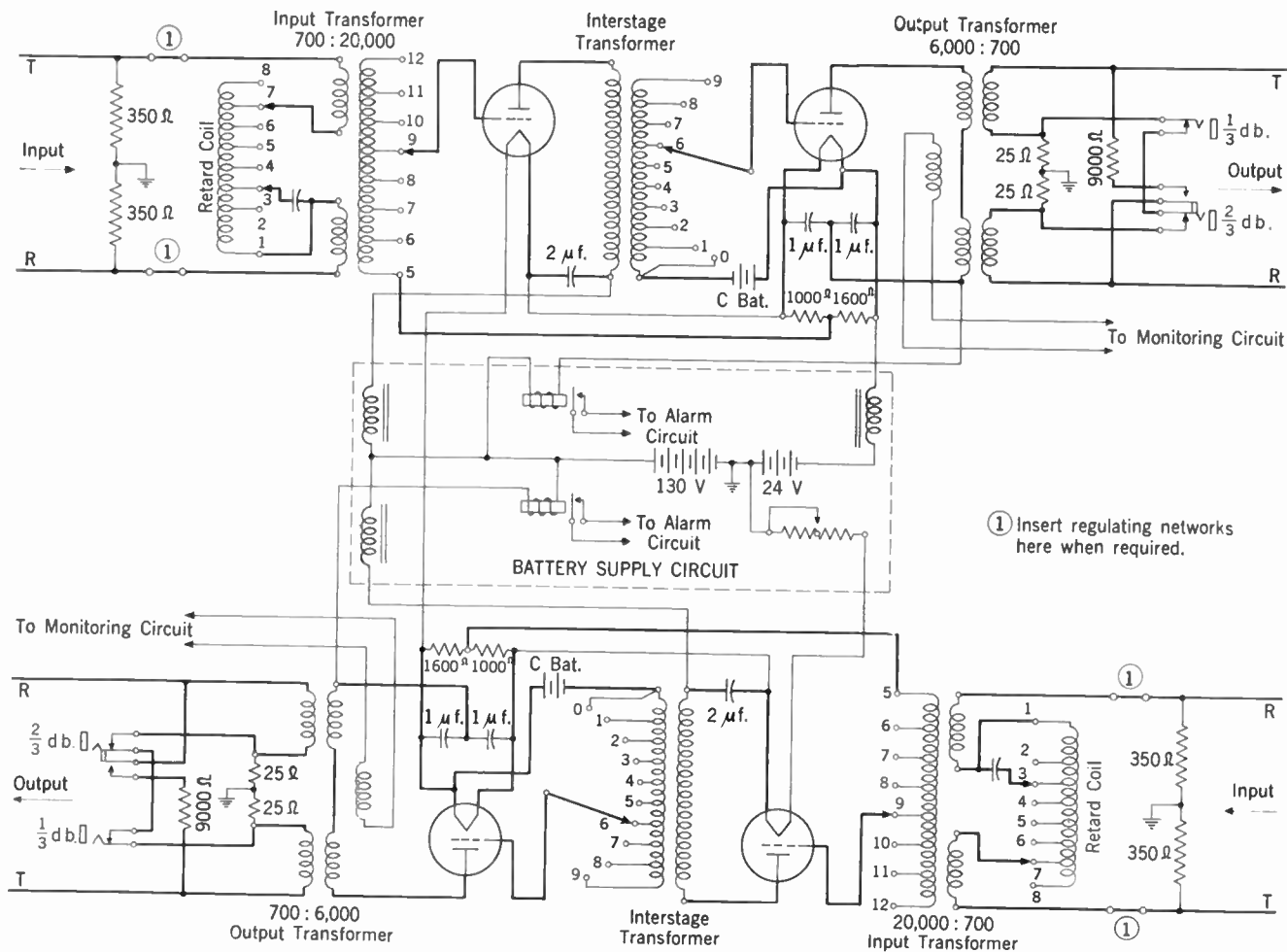
FIG. 25-2 44-TYPE TELEPHONE REPEATER CIRCUIT

overall gain of this repeater is approximately 19.5 db when the potentiometers are on top step, but the gain of the amplifying units themselves must be higher than this to overcome the losses in the hybrid coils and other circuit elements. The gain is essentially flat over the frequency range from about 200 to 3000 cycles.

For 4-wire circuits, the repeater corresponding to the 22-type repeater is known as the **44-type.** In this case, the circuit itself is double-tracked so that there is no necessity for using hybrid coils except at the circuit terminals where the 4-wire circuit is converted to 2-wire for connection to the switchboard. There is therefore no need for line balancing networks at repeater points, and little possibility of a "singing" path around the individual repeater. For this reason, 4-wire repeaters may generally be operated at higher gains than 2-wire repeaters.

The circuit arrangements of the 44-type re-

peater are illustrated schematically by Figure 25-2. It will be noted that each amplifier has two triode stages, the first tube acting as a voltage amplifier and the second as a power amplifier. Transformers are used for interstage coupling, as well as in the input and output circuits. The gain is controlled by adjustable steps on the secondary windings of both the input and interstage transformers. The shape of the gain-frequency characteristic is controlled by an equalizing network connected in series at the mid-point of the primary of the input transformer. The maximum overall gain of this repeater is 42.7 db and is flat to frequencies well above 3000 cycles.

Where extremely stable amplifier operation is required, as for example in the repeaters of telephotograph circuits, the 44-type repeater may be modified for operation with negative feedback. This reduces the maximum overall gain to about 38 db. Feedback is from an output unit made up

[232]

of capacitors and resistors inserted between the second-stage tube and the output transformer, to a similar input unit inserted between the input transformer and the first-stage tube.

Another and more recent design of the voice-frequency telephone repeater is known as the **V-type.** It differs from the 22- and 44-types considerably, both with respect to the amplifiers themselves and the associated equipment arrangements. Hybrid coils, equalizers, filters and regulating networks are associated with the line equipment instead of with the amplifiers, so that the repeater proper consists only of the amplifiers themselves. All repeaters are thus essentially identical and this makes it possible to transfer them freely from one circuit to another, as may be required for maintenance purposes. It also makes possible the use of the same repeaters for
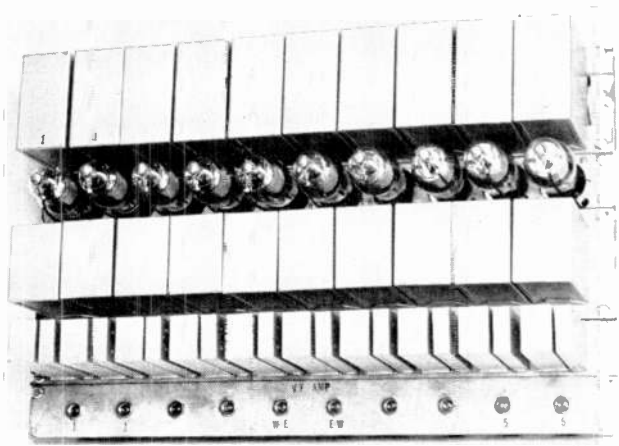
V1 Telephone Repeaters

The miniature technique is used primarily for space saving purposes, and the overall characteristics of the amplifier are approximately the same as discussed in the preceding paragraph, although about 6 db more negative feedback is provided. This gives somewhat greater stability of gain against tube and battery variation. Overall gain may be controlled by adjustment of a potentiometer connected across the secondary winding of the input transformer, which permits a total net variation from a small loss to about a 36 db gain. The vacuum tube used has about twice the transconductance and substantially the same output power as the tube employed in the earlier V-type amplifier. The entire amplifier unit is of the
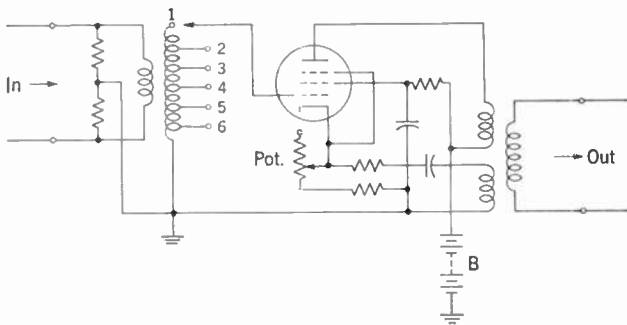
Fig. 25-3 One-half of V-type Telephone Repeater Circuit

either 2-wire or 4-wire operation.

A schematic of one repeater of the V-type is given in Figure 25-3 (showing only one-half). It will be noted that the amplifying element is a pentode rather than a triode. This permits a maximum net gain of about 35 db, even though the feedback circuit causes a reduction of approximately 10 db. Feedback is derived from an extra winding on the output transformer and resistances in the cathode circuit. These latter include a potentiometer, which, together with taps on the secondary winding of the input transformer, serves to control the gain. The gain-frequency characteristic of the amplifier varies less than 1 db over a range of about 250 to 4000 cycles.

Another and later V-type repeater (V3) employs "miniature" pentodes, and correspondingly small equipment parts including input and output transformers, capacitors and potentiometers.
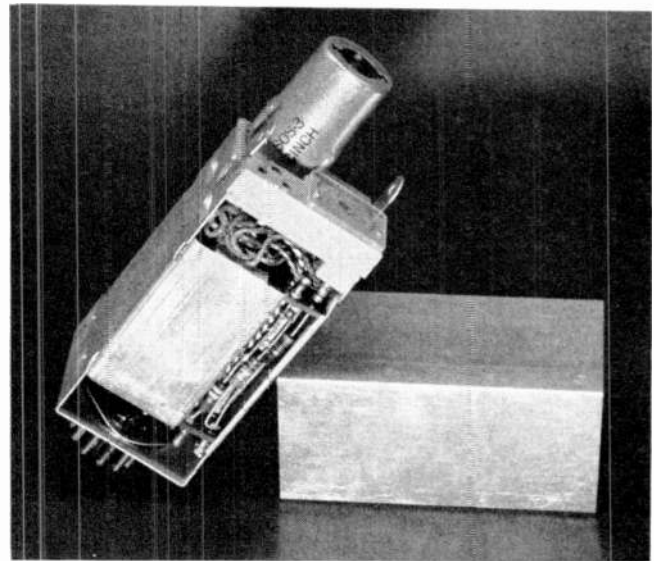
Detail of V3 Repeater Amplifying Unit
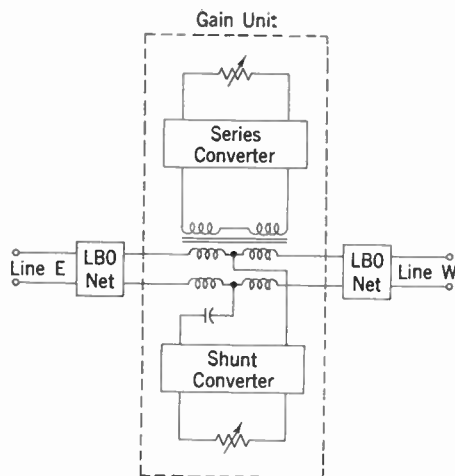
FIG. 25-4 LINE CONNECTION OF E6 REPEATER

plug-in type which provides for quick replacement of defective units and facilitates testing and maintenance.

Where long telephone circuits are used for transmitting program material to the transmitters of a radio broadcasting network, it is frequently necessary to employ amplifiers in the lines at terminal or intermediate points. The "program amplifier" need be only a one-way device, but it must handle a frequency band somewhat wider than is required of the ordinary telephone repeater. One widely used design of such an amplifier employs pentodes connected in two transformer-coupled push-pull stages. Negative feedback is through a gain regulating network providing a maximum net gain of about 42 db. The gain-frequency characteristic is flat from 30 to 8000 cycles and is off by only 1 db at 15000 cycles. Other types of program amplifiers, using R-C interstage coupling, have comparable characteristics.

All of the repeaters discussed thus far are "double-track" devices, employing separate amplifiers for each direction of transmission. A repeater of radically different design is widely used to improve transmission in relatively short circuits such as the inter-office trunks of exchange plant. This is known as a **negative impedance repeater** or **converter** and in the Bell System is coded as an **E-type telephone repeater.** A general schematic of the most recent design of this type of repeater connection (E6) is given in Figure 25-4. Here it may be noted that, instead of being inserted in the line as in the case of other types of repeaters, the gain unit is so connected as not to

break the continuity of the line for the transmission of d-c supervisory signals or other low-frequency signaling currents. This is effected by coupling the series converter through a transformer and by the inclusion of a blocking capacitor in one of the bridged leads to the shunt converter.

The line building out (LBO) networks are included for impedance matching purposes. They are made up of resistors, inductors and capacitors which can be adjusted to fit the electrical characteristics of the particular line with which they are associated so as to present toward the gain unit the impedance of a 900 ohm resistor in series with a 2.14 microfarad capacitor at all frequencies in the voice band.

The converters are dynamic devices which have the ability to convert positive impedances to negative impedances. This is accomplished by
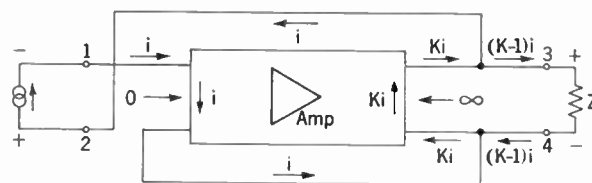
FIGURE 25-5

coupling the output of a special type of amplifier back to its input in an arrangement such as that schematically shown in Figure 25-5. Here the amplifier is one that has input impedance so low as to be practically negligible, very high output impedance, and a current amplification ratio of K. A positive impedance Z is connected across the output at terminals 3-4. The amplifier output is also connected back to the input through the source circuit connected across terminals 1-2.
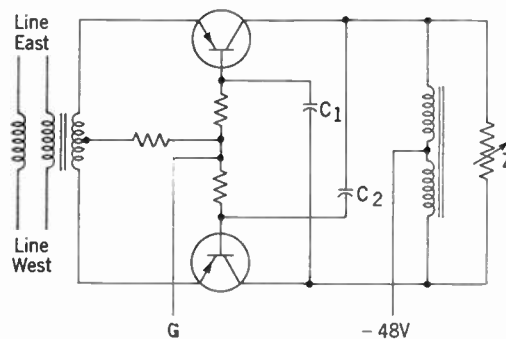
FIG. 25-6 SERIES NEGATIVE IMPEDANCE CONVERTER
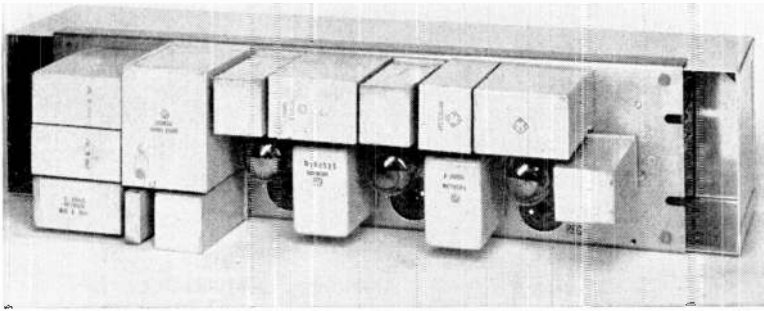
TYPE-K CARRIER LINE AMPLIFIER

Now, if we assume a current, i, flowing to the amplifier input, as indicated by the arrow, the amplifier output current will be Ki and the current through Z will be $Ki\text{-}i$ or $(K\text{-}1)\,i$. The voltage drop across Z is therefore $iZ\,(K\text{-}1)$. But since the amplifier input impedance is nearly zero, this same voltage must exist across terminals 1–2. The impedance at 1–2 is therefore $\dfrac{iZ\,(K\text{-}1)}{i} = Z\,(K\text{-}1)$. The direction of the current and voltage at 1–2 is such, however, that energy is now being fed into the source circuit rather than taken from it. The impedance seen by the source must therefore be considered as negative and its value written as $-Z\,(K\text{-}1)$.

Amplifiers meeting the special requirements specified in the above discussion are readily designed. A push-pull arrangement of two vacuum tube triodes in a grounded grid connection provides a good approximation. A pair of junction transistors in a grounded base connection meets the requirements even more closely. Figure 25-6 shows the essential elements of the series converter of the E6 repeater. The feedback connection here is through the relatively large capacitances C_1 and C_2. The current ratio between emitter and collector is very nearly 1 so that the total current amplification K is approximately 2. The impedance at the converter side of the line transformer is thus almost the exact negative of the positive impedance Z. The amount of negative impedance effectively inserted into the transmission line is therefore directly controlled by the value of Z.

The negative impedance so inserted in the line by the series converter reduces the normal impedance of the line itself and thus permits a corresponding increase in line current for a given applied voltage. The effective reduction in the total line attenuation losses means, of course, that the converter is supplying energy to the line as in the case of other types of repeaters. The amount of net gain that can be realized with the series converter alone is limited in practice by the fact that it tends to introduce a substantial impedance irregularity in the line at the point where it is connected. The severity of this irregularity can be minimized if Z is made up of a rather complex arrangement of resistors, inductors and capacitors which can be connected in various combinations to best adjust to the impedance characteristics of the particular line.

Better impedance matching and consequent higher gains, can be secured by including in the repeater a second converter which is bridged across the center taps of the line transformer, as indicated in Figure 25-4. This "shunt converter" employs somewhat different circuitry than that of Figure 25-6 but operates on the same principle to introduce a negative admittance which, together with the negative impedance of the series converter, presents a net impedance toward the line in both directions of 900 ohms plus $2\mu f$. With this arrangement, the adjustable Z networks in both converters can be made up entirely of resistors
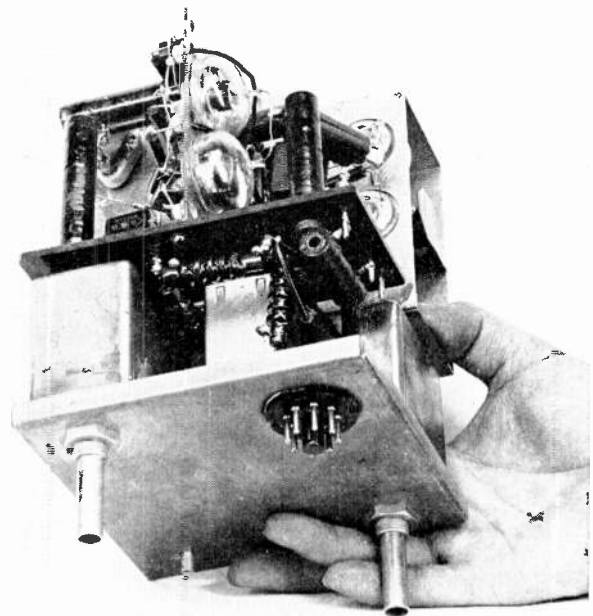
TYPE L1 CARRIER AMPLIFIER

[235]
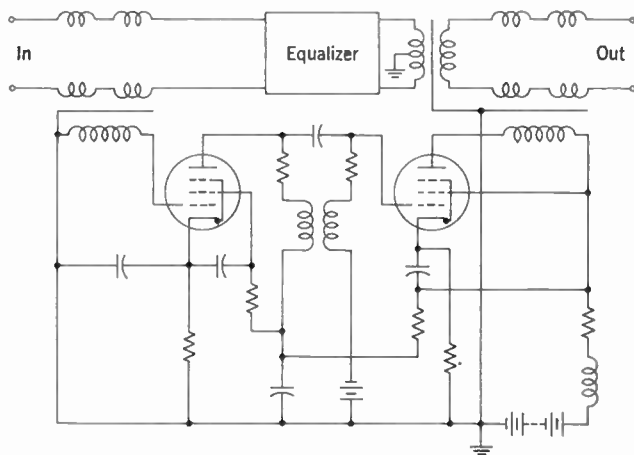
FIG. 25-7 AMPLIFIER USED IN TYPE-C CARRIER SYSTEM

which can be easily adjusted in steps of ½ db (0.1 db if desired) to provide an overall repeater gain of any desired value up to a maximum of about 12 db. Because of these advantages, current designs of E-type repeaters include both series and shunt converters and no provision is made for using either converter by itself.

25.2 Carrier Amplifiers

The several types of carrier systems currently in use in telephone practice employ frequency bands ranging from 4000 cycles up to as high as several million cycles. The amplifiers used in these systems must be designed to handle the entire frequency band of each particular system on a

high fidelity basis with a reasonably flat gain over the total frequency range. All such amplifiers are designed with stabilized feedback circuits.

Figure 25-7 is a schematic of an amplifier used in the type-C carrier system which operates over the frequency range from 5 to 35 kc. As indicated, it consists of two transformer-coupled pentodes with hybrid type input and output transformers. The negative feedback circuit is connected between the two hybrids through an equalizer network circuit. This amplifier has a gain of 50 db, flat from 5 to 35 kc.

Amplifiers for the type-J and type-K carrier systems must handle still broader frequency bands. Thus the type-J amplifier must operate through a range of 36 to 140 kc. It is a three stage amplifier using voltage amplifying pentodes in the first two stages and four power pentodes in parallel in the last stage. Resistance-capacitance coupling is used in the interstage networks. The amplifier has two feedback circuits. The outer feedback circuit extends from the output to the input hybrid transformer through an equalizer network. An auxiliary inner feedback path is connected from the parallel plates of the output tubes to the grid of the input tube through a network which is designed to control the singing margin at frequencies considerably above the normally transmitted band. Type-K carrier amplifiers (12 to 56 kc) also employ three stages, with feedback through a gain adjusting and equalizing network.

Figure 25-8 shows schematically the circuit of
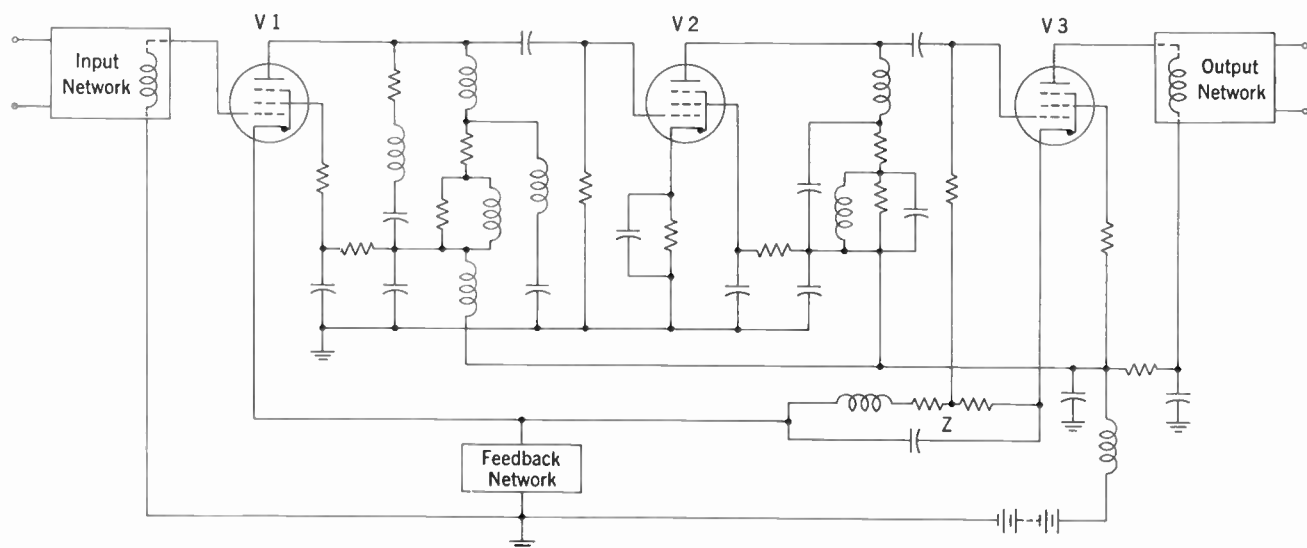
FIG. 25-8 L-CARRIER AMPLIFIER CIRCUIT (PARALLELING TUBES OMITTED)

the amplifier used in type-L1 carrier systems, which operate in the range from 60 to 3000 kc. Although the diagram shows only three tubes, these amplifiers are actually built with paralleling tubes in each stage so that a tube failure will not stop the functioning of the amplifier. Because the amplifier must operate over such a wide frequency range, the interstage networks of this amplifier are quite complex. As a matter of fact, neither of the interstage circuits by itself provides a flat gain, but the two interstages in tandem yield an essentially constant gain between the grid of the first tube and the grid of the last tube over the entire carrier range.

The inter-electrode capacitances of the tubes are minimized by the use of special types of miniature tubes having relatively high transconductance. Two feedback paths are provided, one around the output tube V3, and the other around the entire circuit. The alternating component of the plate current of tube V3 flows from the cathode of that tube through the impedance Z and the feedback network to ground; and thence through the primary of the transformer in the output network to the plate of the tube and back to the cathode. The voltage drop to ground across Z and the feedback network resulting from this current flow is applied to the grid of V3. This local feedback suppresses modulation effects (distortion) developed in the output tube. The alternating voltage developed across the feedback network is applied to the grid of tube V1 through the secondary of the transformer in the input network. This is the main stabilizing feedback of the amplifier circuit. It also provides a means for regulation of the amplifier through appropriate adjustments of the feedback circuit, as is discussed in a later Chapter.

25.3 Broad-Band and Superhigh-Frequency Amplifiers

As was stated in the preceding Chapter, the stages of the usual radio amplifier are coupled by means of single- or double-tuned resonant circuits. This ordinarily permits the satisfactory transmission of a frequency band of the required width. The overall amplification desired in a given case is obtained by using as many stages as may be necessary, with the required output power secured by the use of appropriate power tubes in the final output stage. In most cases, radio amplifiers do not employ stabilized feedback.
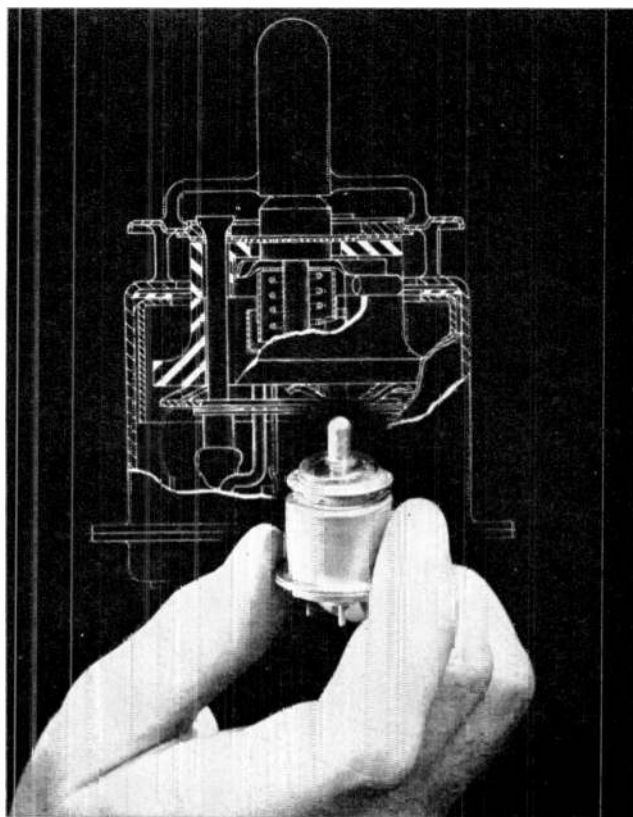
MICROWAVE TRIODE

In cases where very broad-band signals are to be transmitted by radio, such as those of television and L-carrier, reasonably flat amplification over the entire signal range may require the use of a rather large number of stages. For example, the intermediate frequency (IF) main amplifier of the TD-2 radio relay system employs eight stages to give a maximum gain of about 60 db with a bandwidth of about 22 mc. The input circuit, output circuit, and all interstages except the last are double-tuned. The last interstage is triple-tuned. Broad-band IF amplifiers of this type are sometimes "stagger-tuned"; that is, the various interstage networks are tuned to a number of different frequencies within the range of the frequencies to be transmitted.

Amplifiers for use at superhigh-frequencies (3000 to 10,000 mc) present a number of special problems that require rather fundamental design differences. Chief of these perhaps is the fact that the transit time of the electrons in the amplifier tubes becomes a matter of major importance at these extremely high frequencies. As was pointed out in Article 22.4, the adverse effects of transit time likely to be encountered in ordinary vacuum
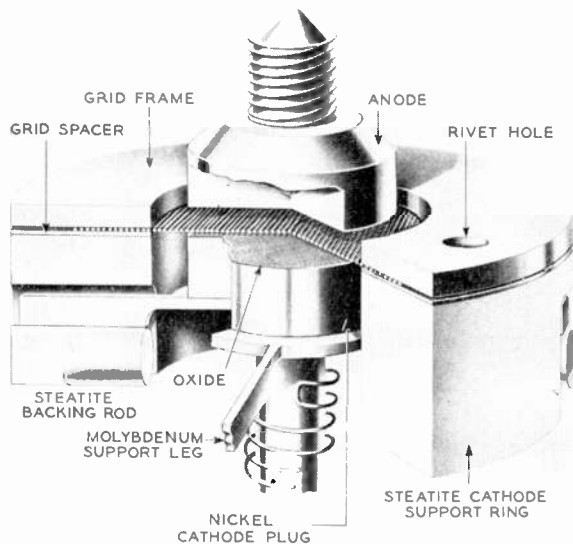
FIGURE 25-9

tubes can be overcome by the use of tubes of the klystron type. These tubes, however, are rather difficult to maintain. Furthermore, present designs of klystrons do not permit the use of more than a limited number of klystron amplifiers in tandem without cumulative noise and distortion becoming excessive.

For application in very long radio relay systems requiring dozens of amplifiers in tandem, accordingly, a special tube (W.E. 416) was developed by Bell Telephone Laboratories. Because of the wide band of frequencies to be handled (at least 20 mc) this tube had to have very high transconductance. To overcome transit time limitations, the elements or electrodes of the tube also had to have extremely close spacing. Fortunately, these two requirements are compatible. The 416 tube is a triode of the so-called planar type, in which the elements are in parallel planes, with the grid grounded to the frame of the structure. A perspective drawing (greatly enlarged) of the elements of the tube is shown in Figure 25-9. The oxide coating of the cathode is .0005" thick. The cathode-grid spacing is .0006". The grid wires are spaced a thousand to the inch and are .0003" in diameter. The plate-grid spacing is .012". The large number of very fine wires employed in the grid structure provides a close approach to a uniform electrostatic shield between cathode and plate, without interfering with the free flow of

electrons. The transconductance of the tube is in the order of 50,000 μmhos, the amplification factor is about 350, and the output resistance 7000 ohms.

This triode always operates in a grounded grid circuit arrangement, where the input is applied between the cathode and ground and the output impedance is between the plate and ground. This automatically eliminates coupling between the output and input circuits through the inter-electrode capacitances of the tube, a feature that is especially useful at very high frequencies where neutralization of internal capacitance coupling by conventional methods is difficult. The possible gain, however, is less than that of the more usual grounded cathode arrangement because of the negative feedback inherent in the fact that the cathode to ground input circuit is included in the plate current path.

Figure 25-10 shows one of these tubes connected in a waveguide "circuit". The input waveguide is coupled through an iris to an input cavity which may be tuned to resonance by a trimming screw across the opening. The grid separates the input cavity from a second resonant cavity of the plate or output circuit. This cavity transforms the plate impedance of the tube to a very low resistance (a fraction of an ohm). A quarter-wavelength

FIGURE 25-10

[238]
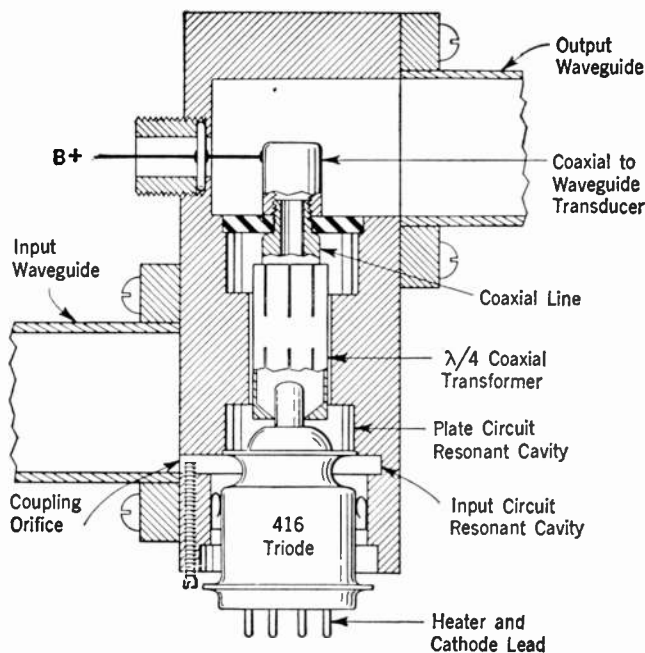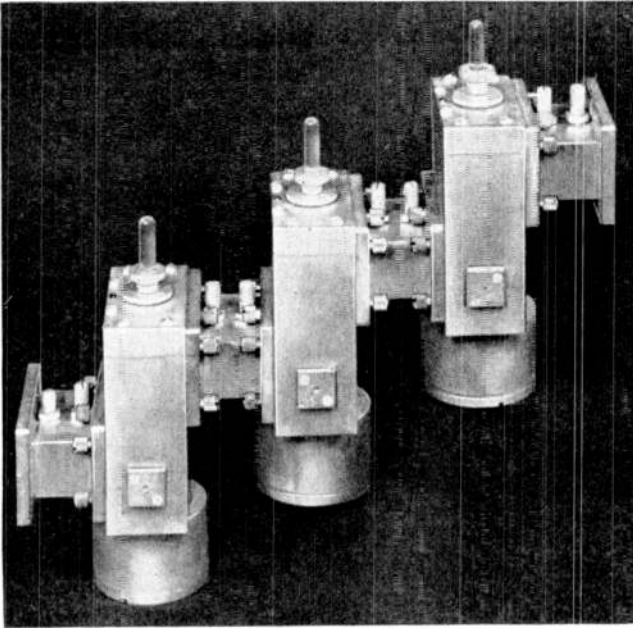
THREE-STAGE MICROWAVE AMPLIFIER

ance to the impedance (about 45 ohms) of a short coaxial line leading to the transducer probe which extends into the output waveguide.

Three tubes of this type are used in the transmitter amplifiers of the TD-2 microwave radio relay system. The three stages of the amplifier, which is illustrated in the accompanying photograph, are connected in cascade through waveguide tuners that effectively form double-tuned critically coupled transformers. The overall gain of the amplifier is normally adjusted to 18 db with an output power of 0.5 watt, although a somewhat higher gain is possible. The overall transmission characteristic is flat over about 20 mc between points 0.1 db down. As the output power of the close-spaced triode is increased, its maximum possible gain decreases. Accordingly, the gains of the three stages of the amplifier are not alike. The first stage has an output of about 80 milliwatts and a gain of about 9 db. The second stage output is about .25 watt with a gain of about 6 db, and the third stage output is about 0.5 watt with a gain of about 3 db.

coaxial line, which can be adjusted vertically to tune the output cavity, matches this low imped-

CHAPTER 26

RECTIFIER, DETECTOR, DISCRIMINATOR, OSCILLATOR AND WAVE-SHAPING CIRCUITS

26.1 Electron Tube Rectifiers

The operation of most electron tubes requires the application of d-c potentials to certain of the tube electrodes—particularly the anodes. These voltages are sometimes supplied by batteries, but it is generally more economical and practical to make use of ordinary commercial power sources when a considerable number of tubes are involved. Since these sources almost invariably supply only a-c power, it is necessary to convert this to d-c power for tube operation. Devices for accomplishing this result are called rectifiers.

Article 7.5 discussed briefly the "barrier-layer" metallic rectifiers that are quite generally used for developing substantial amounts of d-c power in telephone central office power plants. For more limited requirements, such for example as those of the ordinary radio receiver, electron tubes are used to produce the necessary d-c voltage. Rectifying tubes are diodes, and may be of either the high vacuum or gas type. Figure 26-1 shows a diode connected in a circuit arrangement known as a half-wave rectifier. Here the alternating supply voltage is impressed through a transformer across the plate and cathode of the tube. The tube filament is also heated from this same supply voltage by means of an auxiliary transformer winding as shown. Due to the fundamental characteristics of the tube, current can flow only during the time the plate is positive with respect to the cathode. During the positive half of the impressed voltage cycle, accordingly, the tube conducts and the secondary of the transformer is effectively connected to the load. During the negative half of the cycle, no space current can flow and the tube effectively

opens the circuit to the load. The voltage applied to the filter will therefore consist of a series of pulses, as shown in Figure 26-2(A). If the diode is of the high vacuum type, each pulse will have a waveshape practically identical with that of the positive half of the applied a-c voltage. When the applied voltage is a sine wave, as is normally the case, the output current will contain a d-c component equal to the average value over each half-
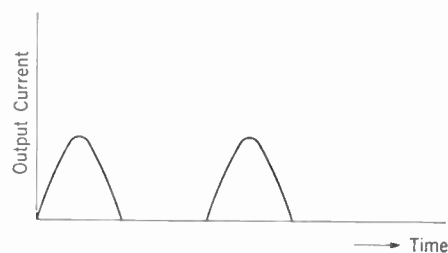
A. Half-Wave Rectifier
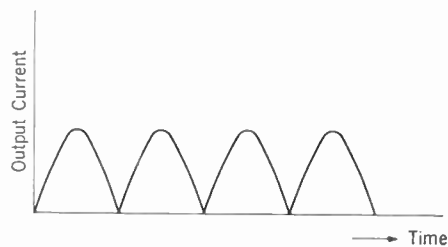
B. Full-Wave Rectifier

FIGURE 26-2

cycle pulse plus a-c components at the frequencies of the input voltage and its even harmonics. These a-c components are not only useless but undesirable since they would tend to introduce noise (power hum) into any circuits to which the rectified voltage might be applied. They are accordingly suppressed by a filter which usually consists of series inductors that present high impedance to the a-c components, and shunt capacitors whose a-c impedance at the frequencies involved is low.

It is fairly obvious that the ratio of the total values of the a-c components to the d-c component

FIG. 26-1 HALF-WAVE RECTIFIER
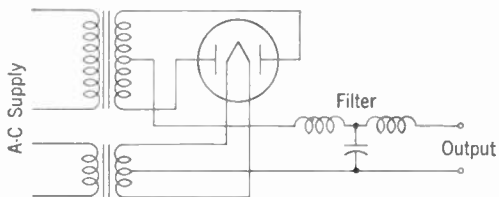
FIG. 26-3 FULL-WAVE RECTIFIER

in the output of the half-wave rectifier is rather large. In other words, this rectifier has a high **ripple factor** and is not very efficient. A better arrangement is the full-wave rectifier circuit shown in Figure 26-3, where the tube has two plates so connected that there will always be current between one of the plates and the cathode. Output current will flow during both halves of the impressed cycle to give a series of positive pulses as indicated in Figure 26-2(B). This output can be analyzed into a d-c component which is obviously twice that of the half-wave rectifier output, plus a-c components having twice the input voltage frequency and its harmonics. The ripple factor is substantially reduced and the d-c output power is four times that of the half-wave rectifier since power is proportional to the square of the doubled current. Because of the lower ripple factor the structure of the smoothing filter may also be somewhat simpler.

Where relatively large amounts of rectified power are needed, as for the operation of ampli-

fiers in radio transmitters or public address systems, gas diodes are frequently employed. Such tubes—especially the mercury-vapor type—are more efficient than high vacuum tubes and will deliver a substantially larger output current for a given size tube. The circuit arrangements for such tubes are no different than those employing high vacuum tubes as shown in Figures 26-1 and 26-3. The output current pulses, however, do not follow the shape of the input so closely because the gas tube does not begin to conduct until the applied voltage has risen to the firing point of the tube.
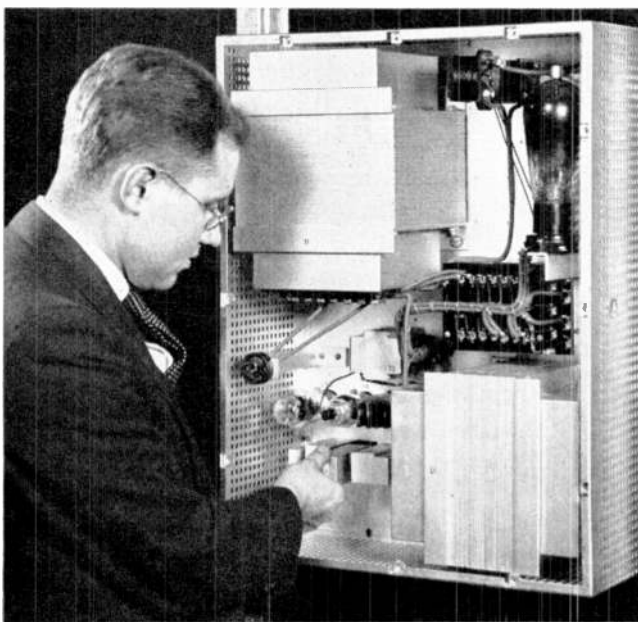
Certain special precautions must be taken in operating a gas tube as a rectifier. Most important of these is the requirement that the cathode be heated to its full operating temperature before
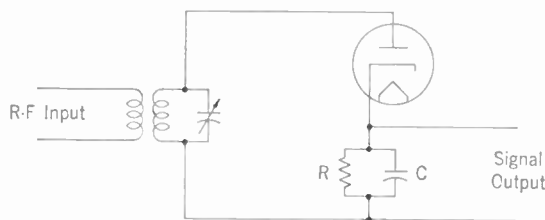
FIG. 26-4 DIODE DETECTOR CIRCUIT

voltage is applied to the plate. This may take several minutes, depending on the particular type tube. It was pointed out in Article 22.1 that a gas tube when operating normally has a small constant voltage-drop between its electrodes, which in the case of the mercury-vapor tube, is only about 14 volts. If voltage is applied to the plate of the tube before the cathode is sufficiently heated to develop its normal emission of electrons and the accompanying normal voltage-drop, the large applied voltage will cause positive ions to move towards the cathode at high velocities. Their impact with the cathode may quickly break down the emitting coating and thus destroy the usefulness of the tube. Neither gas nor high vacuum tubes should be operated at voltages that will cause the plate current to exceed that determined by the maximum cathode emission as this will cause damage through overheating of the electrodes or disintegration of the active cathode material. Care should also be exercised that the maximum inverse voltage rating across the tube is not exceeded. This is particularly true in the case of gas tubes which, having higher conduc-

MERCURY-VAPOR TUBE RECTIFIER

[241]

A. Modulated R-F Wave

B. Output Current of Diode

C. Signal Voltage Across Load Resistance
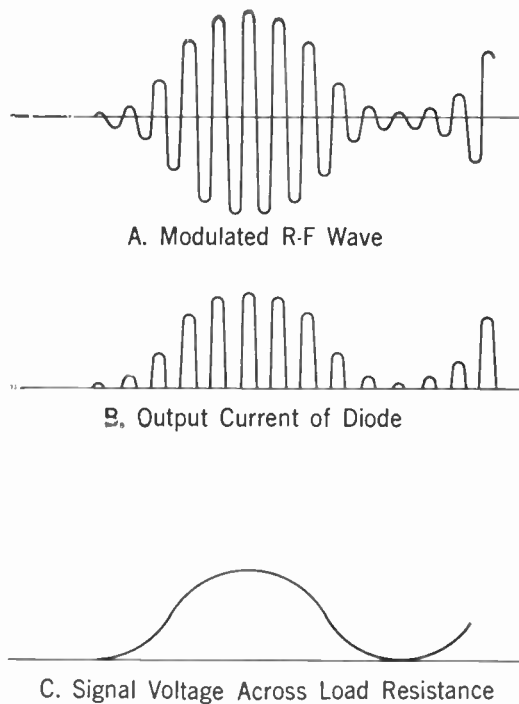
Figure 26-5

tivity, will arc back more easily than the high vacuum tube.

26.2 Detectors

Detector is the name commonly used in radio work to designate the device employed at the receiving end of the circuit to separate the signal from the high-frequency **carrier** wave on which it was super-imposed or **modulated** at the transmitting end of the circuit. Detectors, which may also be properly spoken of as demodulators, are essentially rectifiers that normally handle only very small amounts of power.

The high vacuum diode is an effective detector except at very low levels of radio-frequency input voltage. A simple detector circuit employing a diode is shown in Figure 26-4, and its operation in detecting or separating the signal voltage from the radio-frequency input may be understood by referring to Figure 26-5. Figure 26-5(A) indicates a radio-frequency carrier wave that has been modulated by a lower frequency signal. The amplitudes of the pulses of the radio-frequency wave vary in accordance with the signal which the wave is carrying (see Chapter 27). The tube of course will permit the passage of only the positive half of the cycles, as indicated in Figure

26-5(B). Each positive pulse produces a voltage-drop across the resistor-capacitor circuit R-C proportional to the value of the plate current. Each pulse tends to charge the capacitor, and between pulses some fraction of this charge will leak off through the resistor R. R is of such value, however, that only a very small portion of the charge leaks off before the arrival of the next radio-frequency pulse. As a result, a potential will be developed across the R-C combination which is proportional to the peak values of the r-f pulses, with only a slight r-f ripple component. On the other hand, the time constant of the R-C circuit must be low enough so that the charge on C will be completely dissipated through R between the cycles of the signal wave. Under these conditions, the net voltage-drop across the R-C combination will have the shape shown in Figure 26-5(C), which should be identical with the signal applied to the transmitting circuit. These conditions will be realized in a typical detector circuit when the value of C is in the order of 100 μμf, and R is several hundred thousand ohms.

High vacuum triodes, or multi-electrode tubes, may also be used as detectors. A so-called **plate detector** circuit employing a triode is shown in Figure 26-6. The grid of the tube is biased to the cutoff point so that the tube acts as a rectifier in much the same way as does the diode. The amplitudes of the positive r-f current pulses appearing in the plate circuit follow the amplitude of the signal impressed on the input voltage. Radio frequency is eliminated from the output by means of an appropriate R-C circuit in the same way as previously discussed.
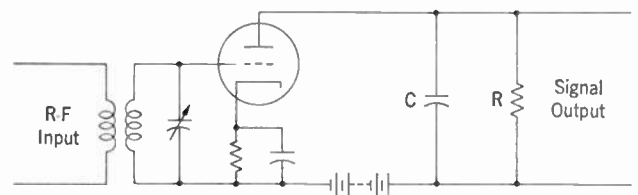
Fig. 26-6 Plate Detector Circuit

Another circuit employing a triode is commonly known as a **grid-leak detector**. This circuit is illustrated in Figure 26-7. It will be noted that the grid of the tube in this circuit is not biased in the usual fashion, and grid current will tend to flow during the positive portion of each r-f cycle. This current, however, is limited by the charge building up in capacitor C_g, which sets up a nega-
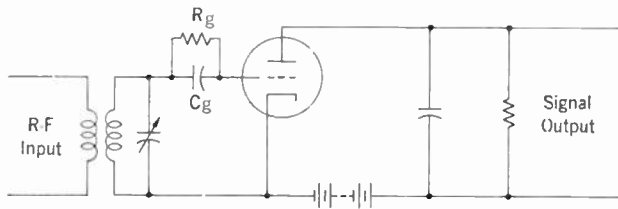
FIG. 26-7 GRID-LEAK DETECTOR

tive potential on the grid that will vary with the
amplitude of the input voltage. The value of R_g
is such that only a negligible portion of the charge
on the capacitor can leak off between r-f pulses,
but the capacitor will be completely discharged in
the time interval between signal pulses. In other
words, the grid-cathode circuit acts exactly like
the diode detector circuit previously discussed.
The net result is that a potential appears on the
grid, which varies in magnitude in accordance
with the impressed signal voltage. These effective
grid bias variations produce corresponding varia-
tions in the plate current of the triode giving an
output voltage across the load resistance which
follows the signal. The circuit thus acts both as
a detector and an amplifier.

Both diode detectors and grid-leak detectors
draw some current from the input circuit, which
means that an effective resistance component is
shunted across the tuned tank circuit. This effec-
tive resistance is high, but it has some tendency
to decrease the sharpness of selectivity of the tank
circuit. The plate detector circuit does not affect
selectivity because the grid never goes positive
with respect to the cathode. On the other hand,
it operates over a somewhat curved portion of
the grid voltage-plate current characteristic, and
therefore causes some distortion of the signal.

At superhigh frequencies, the usefulness of
vacuum tubes as detectors becomes limited be-
cause of their inter-electrode capacitances. De-
tection in these high frequency ranges is usually
accomplished by means of semi-conductor devices
such as the point-contact rectifier or diode dis-
cussed in Article 23.6. Here a sharply-pointed
fine tungsten wire in contact under a small pres-
sure with a wafer of p-type silicon provides a
rectifying junction which is highly reliable, long-
lived and rugged. Because of the very small con-
tact area, it will operate efficiently at extremely
high frequencies (up to 100,000 mc or more).
Typical rectification ratios (ratio of reverse re-
sistance to forward resistance) are in the order

of 100. Higher rectification ratios may be ob-
tained with similar devices employing n-type ger-
manium as the semi-conductor—particularly
when the contact is "formed" to create a tiny
hemisphere of p-germanium directly under the
metal contact point. Germanium semi-conductors,
however, are more sensitive to temperature than
silicon and use of the latter type is preferred
when temperatures above about 60°C. are en-
countered.

A silicon diode having almost ideal characteris-
tics can be made by what is called the "alloyed
junction" method. In this process the contact area
between a wafer of n-type silicon and a fine
aluminum wire is heated until both silicon and
aluminum melt to form a small hemisphere of
alloy which may be considered to be heavily doped
p-type silicon. When permitted to cool, a thin cup-
shaped layer of the p-type silicon is precipitated in
atomic contact with the n-type silicon which forms
the main body of the wafer. This junction will
transmit relatively large currents in the forward
direction while the reverse current is almost negli-
gible and practically independent of the voltage
up to a large critical value where a sudden but
non-destructive breakdown permits free current
flow. The rectification ratio may be as high as
10,000,000.

26.3 Discriminators

The detection of the signal carried by a fre-
quency-modulated wave requires the use of a cir-
cuit that will convert frequency variations into
amplitude variations. One of the more common
circuit arrangements used for this purpose is
known as a **discriminator**. Such a device is also
fundamental to the operation of automatic fre-
quency control circuits.

Figure 26-8 is a simplified schematic of a typi-
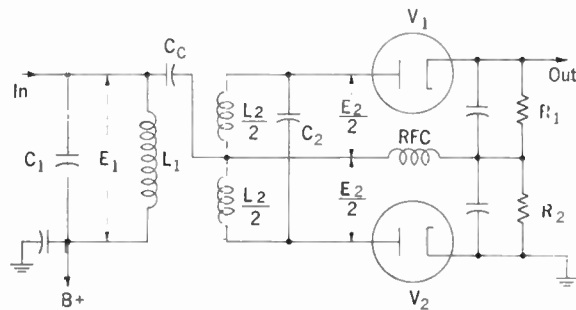
FIG. 26-8 TYPICAL DISCRIMINATOR CIRCUIT

cal discriminator circuit. As will be noted, this circuit employs two diodes connected in a balanced arrangement. The primary and secondary tank circuits L_1C_1 and L_2C_2 are both tuned to the carrier or mean frequency of the r-f input voltage. L_1 and L_2 are also the primary and secondary windings of a transformer. The input voltage across L_1 accordingly induces a voltage in the secondary winding L_2, each end of which is connected to the plate of one of the diodes. With reference to the center tap of the secondary winding, the induced voltages in the two halves of the winding are always equal and opposite in sign; and at the carrier frequency, when the tank circuit impedance is wholly resistive, are 90° out of phase with the input voltage E_1.

In the absence of the coupling capacitor, C_c, the circuit would always apply equal and opposite
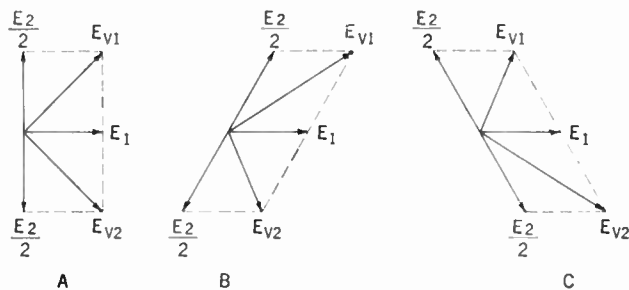
FIG. 26-9 VECTOR DIAGRAM OF VOLTAGES IN DISCRIMINATOR CIRCUIT

voltages $E_2/2$ to the diodes, which would result in the development of equal and opposite voltage drops across the resistances R_1 and R_2, and consequent zero output. The capacitor C_c, however, is relatively large so that the full r-f input voltage E_1 is applied directly to the center tap of L_2. The net voltages applied to the plates of the diodes are accordingly equal to the sum or difference of this voltage and the voltages induced in L_2 by the current flowing in L_1. In other words, the voltage at the plate of V_1 will be $E_1 + E_2/2$, while the corresponding voltage at the plate of V_2 is $E_1 - E_2/2$. When the input voltage is at the carrier or mean frequency to which the tank circuits are tuned, these voltages, E_{v1} and E_{v2}, will be equal as illustrated by the vector diagram in Figure 26-9(A). If the frequency of the input voltage becomes higher than the resonant frequency of the tuned circuits, the inductive reactance of the tank circuits predominates and the phase relationship be-

tween E_1 and E_2 is no longer 90°. Then, as illustrated by Figure 26-9(B), the voltages $E_2/2$ add vectorially with E_1 to give unequal resultant values. This results in a larger voltage drop across R_1 than across R_2 and a consequent positive output voltage with respect to ground. Similarly, if the input frequency is lower than the carrier frequency, the capacitive reactance of the tank circuits predominates and the vector diagram takes the form shown in Figure 26-9(C), resulting in the appearance of a negative voltage in the output circuit. Thus deviations of the input frequency from the mean or carrier frequency produce output voltages corresponding in sign to the direction of the variation and in magnitude to the extent of the deviation.

26.4 Oscillators

In communications work, the term **oscillator** is usually applied to electronic devices which act as generators of a-c sine wave voltages. Practically any vacuum tube or transistor amplifier circuit will function as an oscillator if some part of the output energy is returned or fed back in phase to the input. The minimum requirement for sustained oscillation is that the energy so fed back must be at least as much as the reciprocal of the total amplification. That is to say, for example, if the energy amplification of the circuit is 100 times, at least 1/100 of the output energy must be fed back. This is a condition which it is not at all difficult to obtain. In fact, it is usually necessary in the design of any amplifying circuit to take special precautions to avoid the development of an oscillating condition.

Figure 26-10 represents a simple type of inductively coupled oscillator circuit. Here the amount of energy fed back into the input is determined by the coupling between the coils L_1 and L_2, and the frequency of oscillation is controlled by the values of L_1 and C_1 in the resonant tank circuit.
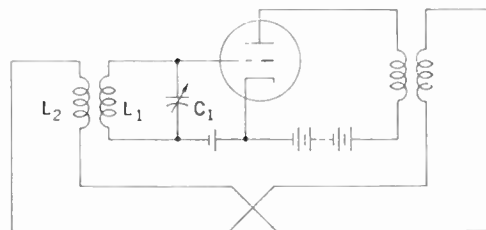
FIG. 26-10 INDUCTIVELY COUPLED OSCILLATOR
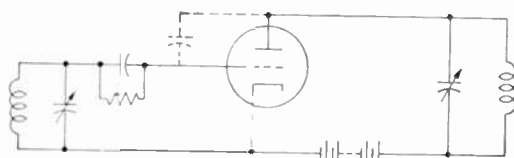
A. Tuned Grid - Tuned Plate Oscillator
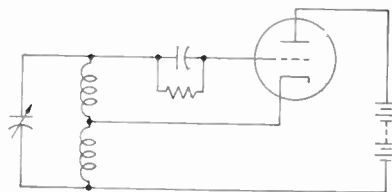
B. Hartley Oscillator

FIGURE 26-11

Many other designs of oscillator circuits are possible. Figure 26-11(A) shows a vacuum tube oscillator circuit in which both the grid and plate circuits include tuned tank circuits. Here the feedback or coupling between the plate and grid is assumed through the inter-electrode capacitance of the tube itself. If this is insufficient, it may be effected by the inclusion of a coupling capacitor as indicated by the dotted lines.

Figure 26-11(B) shows the principle of the well-known Hartley oscillator in which the tuned network is connected between the grid and plate of the tube, and the cathode is connected to a tap-point of the inductor. Thus the current flowing in the plate circuit produces a voltage between grid and cathode whose value depends upon the ratio of the inductive reactances on either side of the tap-point. It may be noted that grid bias is obtained in both circuits of Figure 26-11 by the use of a grid-leak resistor and capacitor. This arrangement has two advantages. It insures that the oscillator will be self-starting, since at the instant that voltage is first applied to the plate the grid will be unbiased and the tube will be working on a high point of its characteristic curve. This will permit an initial surge of current which will begin to charge the grid-leak capacitor and supply sufficient energy to the tank circuit to start it into oscillation. The first few oscillations will continue to build up the charge on the grid-leak capacitor and drive the grid increasingly negative until a steady-state operating condition is reached in which the energy supplied to the tank circuit is just sufficient to overcome its losses and thus maintain oscillations of a constant magnitude. The grid-leak biasing arrangement

also tends to make the oscillator self-regulating because the grid bias will automatically change in accordance with any change in the plate current that may be caused by variations in the load.

The circuits shown in Figures 26-10 and 26-11 do not indicate any load connection, but in any practical case some portion of the plate circuit energy would, of course, be drawn off for application to some other circuit. This can be accomplished by connecting the load directly into the plate circuit, or by connecting it inductively to the coil in the plate tank circuit. It will be evident, however, that such a load connection may affect the constants of the oscillating circuit somewhat, with a possible consequent effect on the oscillating frequency. This may be minimized by inserting a "buffer amplifier" between the oscillator and the load. It may also be avoided by the use of the so-called electron-coupled oscillator shown in Figure 26-12. This circuit employs a tetrode in which the screen grid acts as the plate of an oscillating circuit of the Hartley type. The
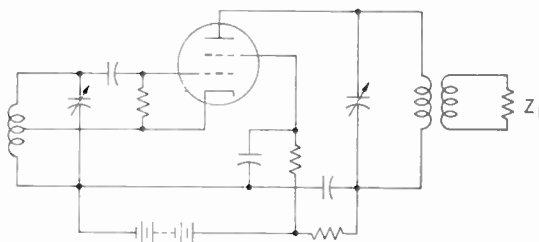
FIG. 26-12 ELECTRON-COUPLED OSCILLATOR

plate circuit couples the oscillating circuit to the load by means of the varying stream of electrons passing through the screen to reach the plate. Changes in the load impedance thus cannot affect the constants of the oscillating circuit itself.

Many applications of oscillators in radio and carrier systems require greater stability of frequency than can be readily obtained with the circuits discussed above. The high degree of stability needed in such cases is usually obtained by employing a piezo-electric crystal in place of the ordinary tank circuit, as indicated in Figure 26-13. As discussed in Chapter 20, such crystals can be cut to precise values of mechanical resonance. When employed in an oscillating circuit as shown, the output frequency will correspond exactly to the resonant frequency of the crystal regardless of variations that may occur in other reactive components of the circuit, or in the char-
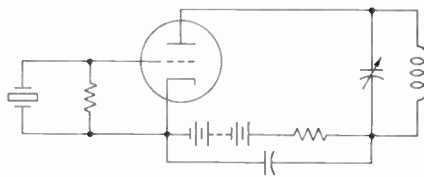
FIG. 26-13 CRYSTAL CONTROLLED OSCILLATOR

acteristics of the tube. If necessary, even greater stability may be secured by enclosing the crystal in a constant temperature oven to preclude any changes in the crystal itself that might result from changes in the ambient temperature.

Crystals can be cut to have natural frequencies ranging from a few kilocycles up to about 15 megacycles. Frequencies much higher than this would require making the crystal too thin for practical use. Stable frequencies above this limiting value may be obtained, however, by using **frequency multipliers** in tandem with a crystal source. These consist essentially of amplifiers operated on a non-linear basis so that their output contains substantial harmonics of the fundamental frequency. In a frequency doubler, the second harmonic appearing in the output of the amplifier is selected by an appropriate tuned circuit, while a frequency tripler would select the third harmonic. Because of their relative weakness, higher harmonics than these are not ordinarily used, but any desired multiplying factor can be obtained by employing as many doubling or tripling stages in tandem as may be necessary. The stability of the end frequency remains as great as that of the originating crystal because the multiplying factor is always a fixed integral number.

Counterparts of the several oscillator circuits

discussed above, employing transistors instead of vacuum tubes, have been designed. This is also generally true of various other vacuum tube circuits mentioned in this and later Chapters. Since the basic circuit principles are essentially the same, however, detailed discussion of the corresponding transistor circuits is omitted for the sake of brevity.

Alternating voltages at superhigh-frequencies can be generated directly by certain types of electron tubes. The klystron tube discussed in Article 22.4 will act as an oscillator if a coaxial feedback connection is made between the catcher cavity and buncher cavity, with both cavities tuned to the
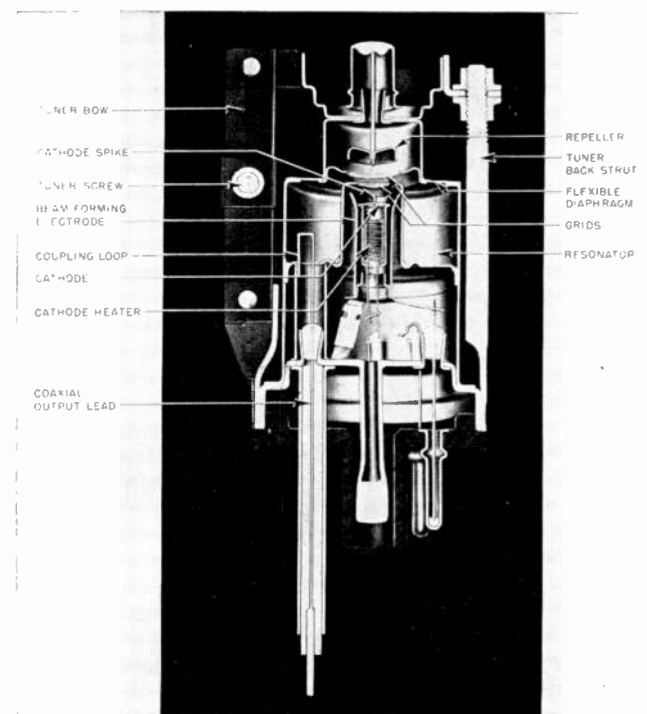
REFLEX-KLYSTRON OSCILLATOR TUBE

same frequency. A somewhat simpler device which will serve the same purpose is the **reflex-klystron** whose basic structure is illustrated in Figure 26-14. This device includes only one pair of grids and one resonant cavity. The cavity acts as both a buncher and catcher because the electrons from the gun pass through its grids twice due to the influence of the "repeller" electrode which carries a high negative charge. As in the ordinary klystron, the electrons proceeding from the gun are velocity modulated in their

FIG. 26-14 REFLEX KLYSTRON

[246]
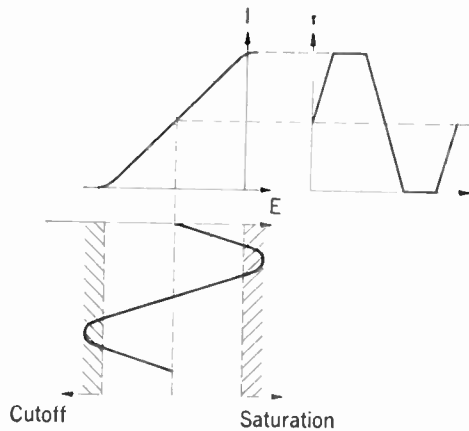
FIGURE 26-15

first passage through the cavity grids. They proceed toward the repeller but are turned back by its negative charge to pass again through the cavity grids in the opposite direction and are finally collected at the accelerating anode of the gun. In the drift interval between their two passages through the cavity grids, the electrons tend to become bunched in accordance with the resonant frequency of the cavity, and when the repeller voltage is appropriately adjusted, the returning electron bunches will pass through the cavity grids in such phase as to give up energy to the cavity.

In addition to its somewhat simpler structure, the reflex-klystron has the advantage over the regular klystron that its output frequency can be varied within limits above or below the natural frequency of the cavity by varying the voltage applied to the repeller. This characteristic is extremely useful in certain frequency modulation systems.

26.5 Wave-Shaping Circuits

There are numerous situations in modern communications practice where it is necessary to alter the shape of voltage waves under controlled conditions. One such situation frequently encountered is the need to limit or clip the peaks of a sine wave. This requirement may apply to either the positive or negative peaks or to both.

Perhaps the simplest method for limiting or clipping both halves of a wave simultaneously is the use of an over-driven amplifier. If the sine wave voltage applied to the grid of a triode is of sufficient magnitude to drive the grid beyond its cutoff point on the negative swing, and above the

saturation point of the plate circuit on the positive swing, the plate current curve will be flattened at its extremities, as illustrated in Figure 26-15. The plate current may thus be made to approach a true square wave as the extent to which the amplifier is over-driven (i.e., as the amplitude of the voltage wave applied to the grid) is increased. Obviously, this same method could be used to limit only the positive or negative pulses by biasing the amplifier grid so that it would be driven above saturation but not below cutoff in one case, and below cutoff but not above saturation in the other case.

In most practical circuits, clipping of the positive half of a wave is accomplished in the grid circuit rather than by driving the plate beyond saturation. If a large resistance is inserted in series with the grid, the applied positive voltage peak will be able to drive the grid only slightly positive because of the large drop across the resistance when grid current starts to flow. The value of the grid voltage is accordingly held close to zero value during the period when the positive input pulse is attempting to drive it positive, and the plate current will therefore have a constant flat value during this same period.

It will be apparent that a square wave generated by the process outlined above will be of much less amplitude than the sine wave from which it was developed. Generally, therefore, the output square wave will have to be amplified if it is to serve any useful purpose. This requires the use of a very broad band video type amplifier circuit, since a square wave contains an infinite number of frequencies. In practice, the bandwidth of such amplifiers may range from 2 to 8 megacycles, depending upon the amount of distortion that can be permitted.

There are various other methods of developing waves having approximately square shapes. One interesting circuit that may be used is the **multivibrator** illustrated in Figure 26-16. This consists
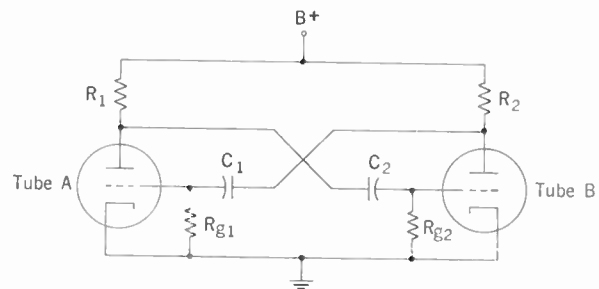
FIG. 26-16 MULTIVIBRATOR CIRCUIT
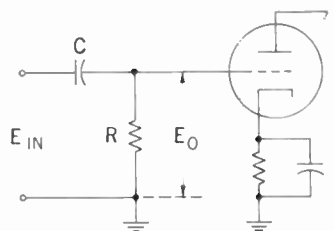
A. R-C Differentiating Circuit
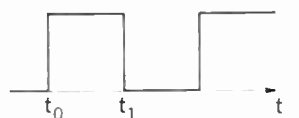
B
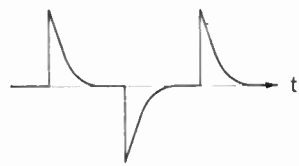
C

FIGURE 26-17

essentially of a two-stage R-C coupled amplifier, with the output of each tube fed back to the grid of the other tube in such phase as to encourage oscillation. The net result is that the tubes become alternately conducting and non-conducting at a frequency rate dependent on the constants of the circuit.

The circuit operation may be followed by assuming an instant when the plate current of one tube is increasing slightly due to some random effect. If this occurs in tube B, for example, it will result in a corresponding decrease of its plate potential. Since this plate is coupled by capacitor C_1 to the grid of tube A, the grid voltage of that tube will also be decreased. This will reduce the current across tube A, which will increase the potential at its plate. The plate of tube A is in turn coupled through capacitor C_2 to the grid of tube B so that this grid's potential is correspondingly increased, resulting in a further increase in the current across tube B. This process continues cumulatively until the current across tube B reaches a maximum value, and the grid voltage of tube A falls below the cutoff value. Tube A will then cease to conduct until enough of the charge on capacitor C_1 has leaked off through the resistor

R_{g1} to permit the grid voltage to rise slightly above the cutoff value. The interval during which tube A is non-conducting is of course determined by the $R_{g1}C_1$ time constant.

As soon as the grid voltage of tube A rises above cutoff, this tube begins to conduct and a cumulative process in the reverse direction proceeds until the current across tube A is at maximum value, and tube B is non-conducting. Although the description of the circuit's action is necessarily somewhat lengthy, the reversing action actually occurs almost instantaneously. The shape of a voltage wave picked off from either of the output resistors R_1 or R_2 is accordingly substantially square. The basic multivibrator circuit discussed is of the so-called "free-running" type, the fundamental frequency of its output wave being determined entirely by its circuit constants. To insure greater stability of operation in practice, an external sine-wave synchronizing voltage of fixed frequency is usually applied to the grid of one of the tubes.

Square waves, as such, have rather limited direct applications in communications systems. They are often used, however, to derive waves of other shapes. Various circuit systems, for example, require uniformly spaced rectangular voltage pulses for their operation. Such pulses may be derived from a square wave by means of a "peaking circuit" known as an **R-C differentiating circuit**. As shown in Figure 26-17 (A), this consists of a simple resistor-capacitor combination working into an amplifier. The circuit derives its name from the fact that the output voltage applied to the grid of the tube is proportional to the time rate of change of the input voltage. Assume that
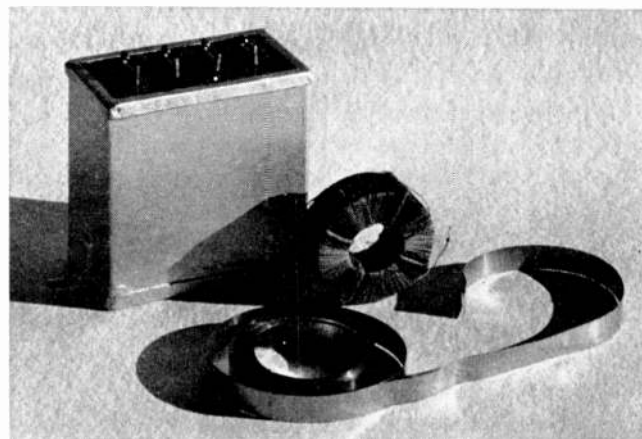
HARMONIC PRODUCING COIL

a square wave voltage, as indicated in the upper diagram of Figure 26-17 (B), is applied to the input. At the instant t_0 the full voltage E is applied across the R-C combination. Since it is impossible for the capacitor to charge instantaneously, the full value of the applied voltage will immediately appear across R, as indicated in the lower diagram of Figure 26-17 (B). The capacitor will then charge in accordance with the time constant RC, which is made short relative to the duration of the applied square pulse. As the capacitor charges, the voltage across R decreases correspondingly until a time equal to about $5RC$ when the capacitor becomes fully charged and the voltage across the resistor is accordingly reduced to zero. At time t_1 the input voltage suddenly drops to zero and the full voltage of the charged capacitor is impressed across R in the opposite direction. As the capacitor discharges to ground,
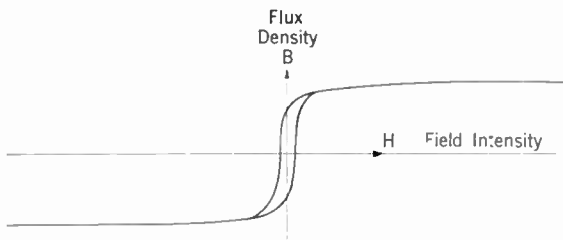
FIG. 26-18 B-H CURVE OF CORE OF SATURABLE INDUCTOR

the voltage across R gradually reduces to zero, as shown.

In most applications of peaked waves only pulses of one polarity are desired. In this event, the output wave, indicated in the lower diagram of Figure 26-17 (B), may be applied to a clipping circuit that will remove the negative pulses entirely. The sharp tops of the positive peaks may also be clipped as required to leave a series of narrow rectangular pulses, as shown in Figure 26-17 (C).

At relatively low frequencies, peaked waveshapes may be developed by entirely different methods employing devices known as **saturable inductors**. These are small coils wound on a core of permalloy ribbon as illustrated in the accompanying photograph. As indicated by the B-H curve of Figure 26-18, the core material of the coil is of such nature that the magnetic field passes from negative to positive saturation with a com-
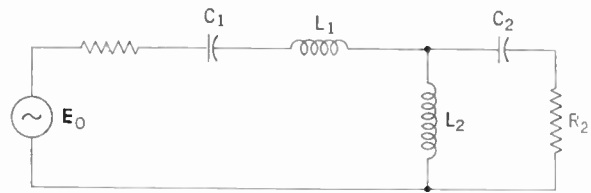
FIGURE 26-19

paratively small change in the field intensity. In other words, the coil becomes saturated very quickly and with relatively little current in its windings. This means that the inductance of the coil is high at low current values, but nearly zero as soon as the saturation point is reached. For use in a peaking circuit, the coil is connected at L_2 as shown in Figure 26-19, where the series circuit consisting of C_1, L_1 and L_2 is made approximately resonant.

The behavior of the circuit may be understood by analyzing what happens during a single cycle of an applied sine wave voltage E_0. This applied voltage is shown in the usual manner in Figure 26-20 (A). As this voltage increases from zero, the inductance of the coil bridged across the line will at first be high, and as a result current will flow into the capacitor and the load R_2. This current is pictured by the small section ab of the curve of Figure 26-20 (B). When the applied voltage increases to the critical value, X, however, the core of the coil becomes saturated and the inductance of the coil immediately decreases to zero.

FIGURE 26-20

[249]
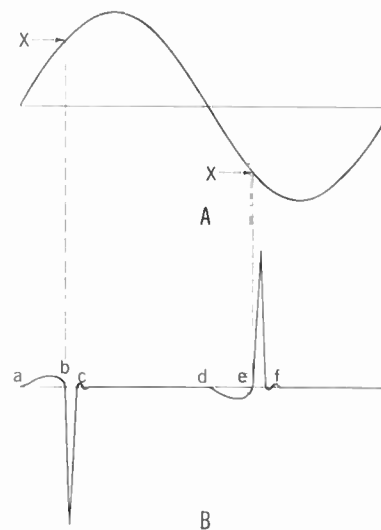
As the coil has quite low resistance, it then becomes effectively a short across the line and no additional current flows from the generator into the load. On the contrary, the charged capacitor C_2 discharges through the coil, causing the sharply peaked negative current surge shown in the section bc. For the remaining part of the positive pulse of the applied voltage, the coil continues to act as a short-circuit, the capacitor remains discharged, and no current flows in the load. When the applied voltage reverses in direction, the coil again presents a high inductance to the low values of negative voltage applied and a small negative current, de, flows into the capacitor and the load.
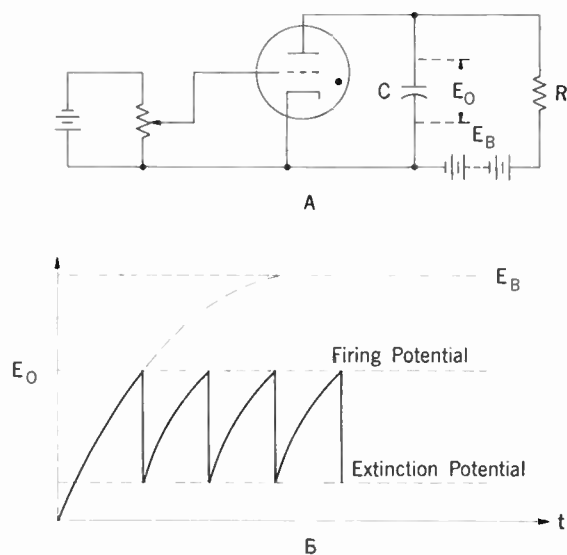
FIG. 26-21 GAS TRIODE SAW-TOOTH GENERATOR

But again, as soon as the coil becomes saturated, the capacitor discharges to cause the sharp positive peak of current ef.

Another wave shape that is used extensively in the operation of cathode-ray oscilloscopes and television cameras and receivers is the so-called saw-tooth wave. This, ideally, is a wave that rises linearly to a specified maximum value in a definite time interval and then returns very rapidly to its original value. One widely used circuit arrangement for generating such a wave employs a gas triode connected as indicated in Figure 26-21 (A). As the B battery circuit is closed, the capacitor C charges through the resistance R at a rate determined by the RC time constant. Until the firing potential of the tube is reached, the voltage

E_0 across the capacitor will build up in accordance with the solid line curve of Figure 26-21 (B). At the firing potential, the value of which is controlled by the grid bias voltage, the tube suddenly becomes conducting and establishes a low resistance shunt path across the capacitor. The capacitor then discharges quickly through the tube until its voltage drops to the extinction potential when the process repeats itself. It will be seen that the time length of each saw-tooth pulse is determined by the RC time constant and the firing potential of the tube. The frequency rate can be stabilized by applying to the grid from an external source a small sinusoidal synchronizing voltage having a frequency slightly higher than the natural period of the tube circuit. This causes the grid bias to vary in accordance with the applied synchronizing voltage so that the tube fires at the same specific point in each sine-wave cycle.

It may be noted from Figure 26-21 (B), that the voltage rise across the capacitor C is not precisely linear since a capacitor charges according to an exponential curve rather than a straight line. However, if the B battery value is made large with respect to the tube firing potential so that the tube will break down when the voltage across the capacitor has risen to only about 5% of the B battery voltage, the rising saw-tooth voltage curve
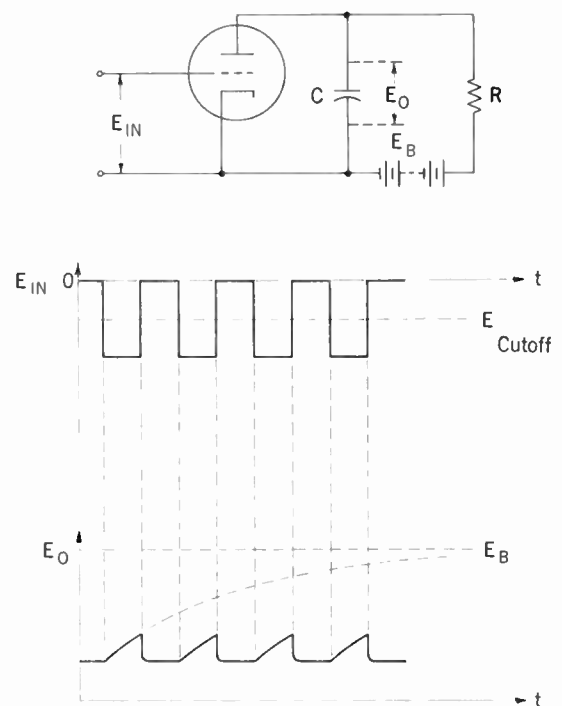
FIG. 26-22 TRIODE SAW-TOOTH GENERATOR

will be sufficiently linear for practical purposes.

Saw-tooth voltage waves may also be developed from a circuit employing a high vacuum triode, to the grid of which is applied a uniform series of negative rectangular pulses as indicated in Figure 26-22. In this arrangement, the pulses applied to the grid have sufficient amplitude to drive the grid well beyond cutoff. During the cutoff period, which is determined by the width of the applied pulse, the capacitor C charges to a maximum value controlled by the RC time constant. When the grid bias is reduced to zero by the trailing edge of the pulse, the tube becomes conductive and the capacitor is quickly discharged. The output voltage E_0 accordingly drops to the normal value established by the B battery at zero grid bias and remains there until the next input pulse is applied to the grid.

CHAPTER 27

MODULATION

27.1 The Carrier Principle

In radio communication systems, and in the various types of carrier systems, transmission of signals is effected by impressing the signal voltage on a carrier wave having a relatively high frequency. The signal is thus transmitted to its destination by electrical waves whose frequencies are normally more nearly comparable in value to the carrier frequency than to the signal frequency.

A. Impressed Telegraph Signal

B. Unmodulated Carrier

C. Carrier After Modulation By A.
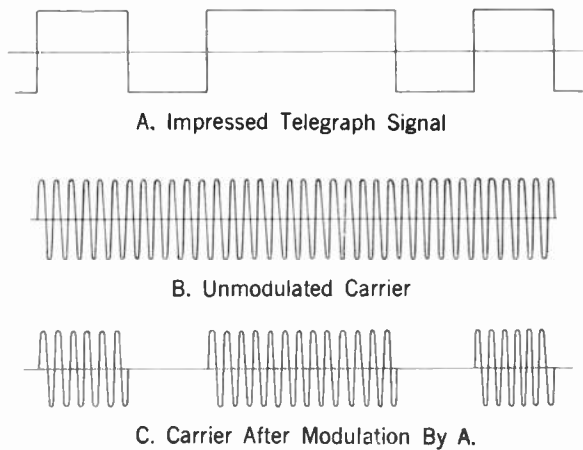
FIG. 27-1 MODULATION IN TELEGRAPH SYSTEMS

quency. The basic reason for this procedure in both carrier and radio systems is to make possible the transmission of a number of different signals over the same transmitting medium without mutual interference by placing each signal in a different portion of the frequency spectrum. In radio transmission, such a procedure is also necessitated by the fact that efficient electromagnetic radiation in space can only be attained at high frequencies. It should be noted, however, that the total width of the transmitted carrier frequency band cannot be less than the sum of the bandwidths of all the signals carried—whether the signals are only a few cycles wide as in telegraph, or millions of cycles wide as in television.

The process of impressing the signal on a carrier is known generally as **modulation.** The inverse process, whereby the signal is retrieved from the modulated carrier-wave, is usually called demodulation in carrier systems and detection in radio systems. There are various methods of modulating carrier waves so that they will effectively transmit signals. The most commonly used of these at the present time are amplitude modulation and frequency modulation.

27.2 Amplitude Modulation

In amplitude modulation, the amplitude of the carrier wave is varied in accordance with the variations of the signal wave. The degree of difficulty involved in modulation depends upon the nature of the signal. For a telegraph signal such as that shown in Figure 27-1 (A), the method is very simple and consists merely in interrupting the supply of carrier frequency to the line during negative pulses of the telegraph signal and permitting it to flow during positive pulses. The result is a series of "spurts" of current at the frequency of the particular carrier channel, as indicated in Figure 27-1 (C).

A. Unmodulated Carrier

B. Voice Signal

C. Carrier After Modulation By B.
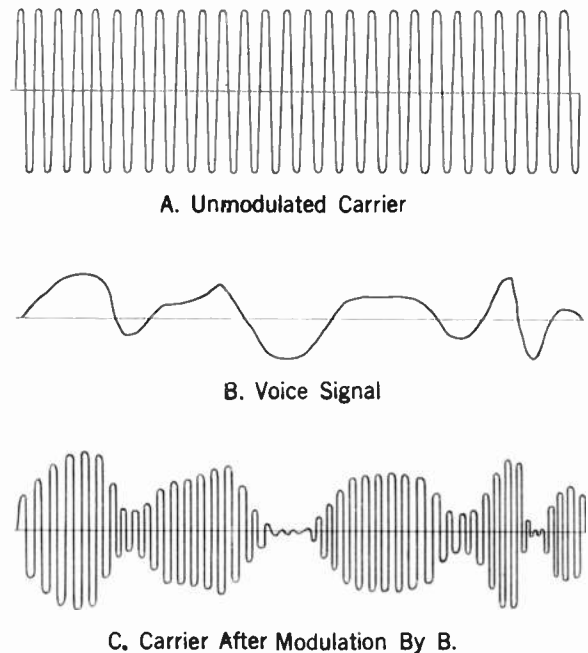
FIG. 27-2 MODULATION IN TELEPHONE SYSTEMS
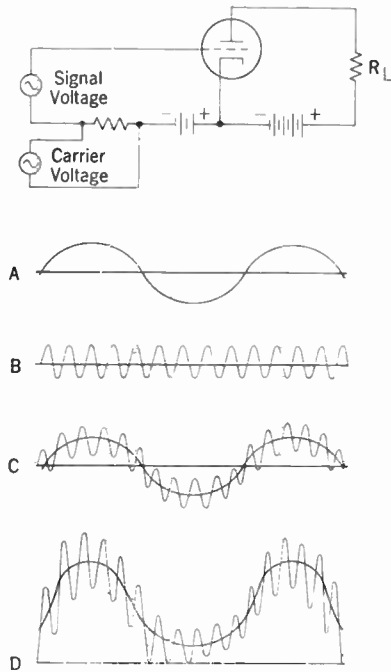
FIG. 27-3 CURRENTS IN MODULATOR CIRCUIT

In telephony, since the variations in voice current are much more complex than telegraph current, the process is more involved. This is indicated by Figure 27-2 where A is a representation of the unmodulated carrier, B is a representative voice signal, and C is the modulated carrier. It will be noted that the outline or "envelope" of the modulated wave has the form of the voice signal wave. This effect is not different in principle from the action of an ordinary telephone transmitter, where the direct current supplied by the local or central office battery is varied or modulated by the sound waves of the voice impinging on the transmitter button. The output current from the transmitter is then a varying direct current consisting of the initial unvarying battery current, with the changing voice current superimposed upon it.

The wave of Figure 27-2 (C) could be obtained by connecting a transmitter in series with the carrier generator, just as the battery is in series with the transmitter in the ordinary subset. The disadvantages of such a scheme will be apparent, however, and in practice vacuum tubes or other kinds of rectifying devices are used for this purpose.

When vacuum tubes are used as modulators in carrier telephone systems, the modulating effect is ordinarily obtained by applying a biasing voltage to the control grid of the tube of such magnitude that the tube operates on the definitely curved part of its grid voltage-plate current characteristic. Under these conditions, the amplification supplied by the tube is not constant but varies with the value of any changing voltage applied to the grid.

In the simple circuit of Figure 27-3, assume that a voice voltage as represented by A is supplied to the circuit together with the carrier voltage represented by B. For simplicity the voice voltage is here assumed to be sinusoidal in form although this, of course, is not usually the case. The two voltages, being in series, add together to give the voltage represented by C impressed on the grid of the tube. Now if the fixed grid bias is given the value E_c indicated by Figure 27-4, and the characteristic curve of the tube is as there shown, the impressed grid voltage will cause a plate current of the form shown in Figure 27-3 (D).

It would be difficult to determine from observation what frequencies make up this rather complex wave, although it is reasonably evident that it must include more than the two frequencies that were applied at the grid. However, mathematical analysis of the plate current wave is relatively straightforward if we assume what is a close approximation to the truth—that the characteristic curve of the tube is parabolic in form over the
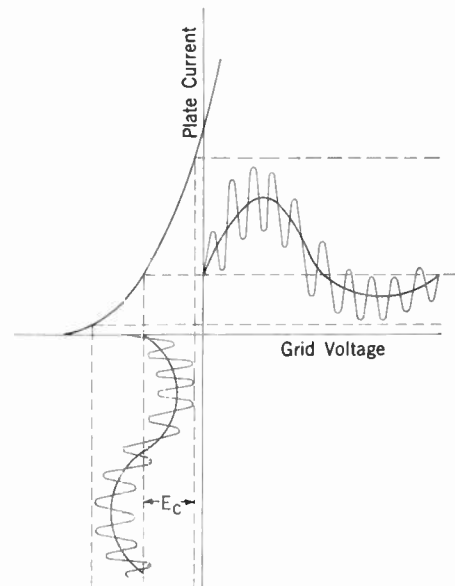
FIG. 27-4 VACUUM TUBE AS MODULATOR

| 253 |

operating range employed. In this event the relationship between plate current and grid voltage may be written as a simple quadratic equation, thus:

$$i_b = K(E_b + \mu E_c + \mu e)^2 \qquad (27{:}1)$$

where—

i_b = instantaneous plate current

K = a constant

E_b = plate battery potential

μ = voltage amplification factor of the tube

E_c = "C" battery or control grid biasing potential

e = instantaneous alternating potential applied to the control grid.

All of these values may be assumed to be held constant during the operation of the tube except-ing i_b and e. Expanding the equation, we have—

$$i_b = K[(E_b + \mu E_c)^2 + 2(E_b + \mu E_c)\mu e + \mu^2 e^2]$$

or, writing a_1 and a_2 for the coefficients of e and e^2 respectively,

$$i_b = K(E_b + \mu E_c)^2 + a_1 e + a_2 e^2 \qquad (27{:}2)$$

where $a_1 = 2K\mu(E_b + \mu E_c)$

and $a_2 = K\mu^2$.

Since the impressed voice and carrier currents represented in Figures 27-3 (A) and (B) are both sinusoidal in form, they may be indicated mathematically by sine functions of time as $A \sin Vt$ and $B \sin Ct$ respectively, where A and B are constants. The applied input voltage, e, is then—

$$e = A \sin Vt + B \sin Ct. \qquad (27{:}3)$$

Substituting (27:3) in equation (27:2), we have for the output current—

$$i_b = K(E_b + \mu E_c)^2 + a_1(A \sin Vt + B \sin Ct) + a_2(A \sin Vt + B \sin Ct)^2$$

and, expanding—

$$i_b = K(E_b + \mu E_c)^2 + a_1 A \sin Vt + a_1 B \sin Ct + a_2 A^2 \sin^2 Vt + 2a_2 AB \sin Ct \sin Vt + a_2 B^2 \sin^2 Ct.$$

$$(27{:}4)$$

Making use of the trigonometric relationships—

$$\sin^2\theta = \tfrac{1}{2} - \tfrac{1}{2}\cos 2\theta$$

and

$$\sin\theta \sin\phi = \tfrac{1}{2}\cos(\theta - \phi) - \tfrac{1}{2}\cos(\theta + \phi)$$

we may expand further to obtain—

$$\begin{aligned} i_b &= K(E_b + \mu E_c)^2 + a_1 A \sin Vt + a_1 B \sin Ct \\ &\quad + \tfrac{1}{2}a_2 A^2 - \tfrac{1}{2}a_2 A^2 \cos 2Vt + a_2 AB \cos(C - V)t \\ &\quad - a_2 AB \cos(C + V)t + \tfrac{1}{2}a_2 B^2 - \tfrac{1}{2}a_2 B^2 \cos 2Ct \\ &= K(E_b + \mu E_c)^2 + \tfrac{1}{2}a_2(A^2 + B^2) + a_1 A \sin Vt \\ &\quad + a_1 B \sin Ct - \tfrac{1}{2}a_2 A^2 \cos 2Vt - \tfrac{1}{2}a_2 B^2 \cos 2Ct \\ &\quad + a_2 AB \cos(C - V)t - a_2 AB \cos(C + V)t. \end{aligned}$$

$$(27{:}5)$$

The first and second terms of this equation are constants representing direct current which, of course, will not appear on the line side of the output transformer. The third and fourth terms are merely amplified currents of voice and carrier frequency respectively; the fifth and sixth are currents of double these frequencies; and the last two represent respectively the difference and the sum of the carrier and voice frequencies. If the voice and carrier frequencies applied to the grid had been, for example, 1000 and 10,000 cycles respectively, the output of the circuit would have contained currents of frequencies 1000, 10,000, 2000, 20,000, 9000 and 11,000 cycles. Practically, of course, applied voice currents would contain numerous frequencies which might have any values between, say, 200 and 2700 cycles, and the output current would vary accordingly. Thus, the output frequency indicated in equation (27:5) as the sum of the voice and carrier frequencies, might occupy any value in the band of frequencies between $(C + 200)$ and $(C + 2700)$.

These sum and difference frequencies are called the upper and lower modulation components, respectively, or, more commonly, the **upper and lower side-bands, and either one of them is by itself capable of carrying the signal current to the receiving end of the circuit.** In carrier telephone practice, it is customary in many cases to suppress by means of filters or otherwise, all of the frequencies in the output of the modulator except one side-band for transmission over the line.

It is obviously desirable also to so arrange the modulator circuit that the current to be transmitted over the line has the largest possible value, and the currents that are not needed have relative small values, thus making feasible the utilization of the greatest possible part of the modulator tube's output energy. This result can be to a degree achieved by properly adjusting the values of the constants a_1, a_2, A and B in equation (27:5). Referring to this, it will be noted that

[254]

if a_1 is made very small, the voice and carrier frequencies may be practically eliminated from the output. This may be accomplished within limits by giving E_c a large negative value, in which case the factor $(E_b + \mu E_c)$, in the expression $(a_1 = 2K\mu |E_b + \mu E_c|)$ may be made to approach zero, reducing a_1 correspondingly.

In carrier systems employing vacuum tube modulator circuits, the method generally employed to control both the absolute and relative magnitudes of the output components depends on the use of a balanced tube arrangement. Such a modulator circuit is shown schematically in Figure 27-5. Under ideal conditions, the output of this circuit includes only the voice frequency and the two side-bands. The carrier frequency itself, as well as harmonics of either voice or carrier are automatically suppressed.

The two tubes of the balanced circuit are so arranged that one-half of the voice voltage, e_v, will be applied to the grid of each tube. But, with transformer connections as shown in the diagram, the voice voltage applied to the grid of tube 1 will be positive with respect to ground at the same time that the voice voltage applied to the grid of tube 2 is negative, and vice versa. The carrier voltage, e_c, on the other hand, is applied in series with the common grid biasing voltage so that its
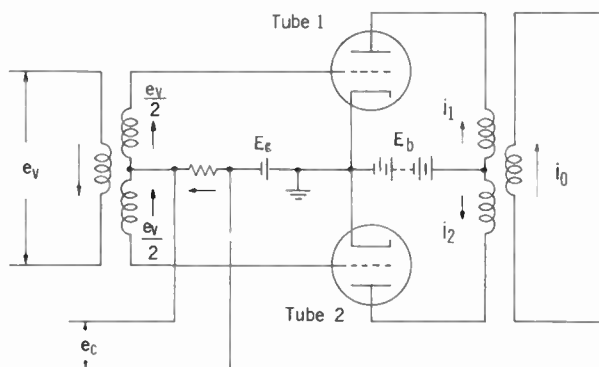
Fig. 27-5 Balanced Tube Modulator Circuit

value and sign is the same on the grids of both tubes. The net alternating voltage applied to the grids of the tubes at any instant accordingly has the following values:

$$e_1 = e_c + \frac{e_v}{2}$$

$$e_2 = e_c - \frac{e_v}{2}$$

These values of input voltage may be substituted in equation (27:2) to determine the plate current of each tube. This gives, for tube 1—

$$i_1 = K(E_b + \mu E_c)^2 + a_1\left(e_c + \frac{e_v}{2}\right) + a_2\left(e_c + \frac{e_v}{2}\right)^2$$

and, for tube 2—

$$i_2 = K(E_b + \mu E_c)^2 + a_1\left(e_c - \frac{e_v}{2}\right) + a_2\left(e_c - \frac{e_v}{2}\right)^2$$

These currents, it will be noted, flow in opposite directions in the primary winding of the output transformer. Therefore, their difference determines the value of the current in the secondary. Since no direct current passes through the transformer, the secondary current may accordingly be written as—

$$i_0 = i_1 - i_2 = a_1 e_c + \frac{a_1}{2}\ e_v + a_2 e_c^2 + a_2 e_c e_v + \frac{a_2}{4}\ e_v^2$$

$$- a_1 e_c + \frac{a_1}{2}\ e_v - a_2 e_c^2 + a_2 e_c e_v - \frac{a_2}{4}\ e_v^2$$

$$= a_1 e_v + 2a_2 e_c e_v \qquad (27:6)$$

Then, by substituting $A \sin Vt$ for e_v and $B \sin Ct$ for e_c, we have—

$$i_0 = a_1 A \sin Vt + 2a_2 A B \sin Ct \sin Vt \quad (27:7)$$

By the second of the trigonometric relationships cited above, equation (27:7) may be converted to read—

$$i_0 = a_1 A \sin Vt + a_2 A B \cos (C - V)t$$
$$- a_2 A B \cos (C + V)t \ (27:8)$$

The three terms of the above equation represent the voice frequency and the two side-bands of the carrier frequency. As in the single tube modulator circuit discussed earlier, the voice-frequency term may be kept small in comparison to the side-band terms by making a_1 as low in value as practicable. Since there is no carrier-frequency term in this equation, an even more effective method of insuring that the greater part of the output energy shall be represented by the side-band terms is to make B much larger than A— that is, to make the amplitude of the applied carrier voltage much larger than that of the applied voice voltage.

It should be noted that in all of the above it has been assumed that the characteristic curve of the modulator tubes had the ideal parabolic form. This is only approximately true in practice and in so far as the curve departs from this ideal, fre-

[255]

quency components additional to those indicated in the above mathematical expressions will appear in the output. Further, a single frequency has been assumed for the applied signal voltage in all cases. Actually, a voice signal normally includes several different frequencies which are applied simultaneously to the modulator input. There will be a certain amount of inter-modulation between these signal frequencies and some of the resultant harmonics or sum and difference components may have frequencies within the range of the useful side-band, thus tending to cause distortion. However, it may be seen from both equations (27:5) and (27:8) that the magnitude of these disturbing frequencies will be proportional in all cases to A, the amplitude of the applied signal voltage. By making the carrier voltage much larger than the signal voltage, accordingly, these frequencies may be kept low enough in value so that their distorting effect is practically negligible.

Demodulation in carrier systems is effected by circuits that are essentially the same as the circuits used for modulation and operate on the same principle. Thus, when the upper side-band, $C + V$, is received over the transmission line, it may be applied to the grid of a vacuum tube together with a carrier C, identical in frequency to the carrier employed at the sending end. The total alternating voltage applied to the grid may then be written as—

$$e = A \sin (C + V)t + B \sin Ct$$

If this value is substituted for e in equation (27:2) and expanded, the resultant output plate current will be found to include—

    C—The carrier frequency.

  $(C + V)$—The impressed side-band frequency.

    $2C$—Twice the carrier frequency.

 $2(C + V)$—Twice the impressed side-band frequency.

$(C + V) + C = 2C + V$—The sum of carrier and side-band.

$(C + V) - C = V$—The difference of carrier and side-band, which is the **voice frequency.**

Similarly, in the case of the balanced tube circuit employed as a demodulator, if the upper side-band $(C + V)$ is applied, the output frequencies will be V, $C + V$, and $2C + V$. This will be apparent from equation (27:8) where $(C + V)$ may be substituted for V to determine the demodulation products.
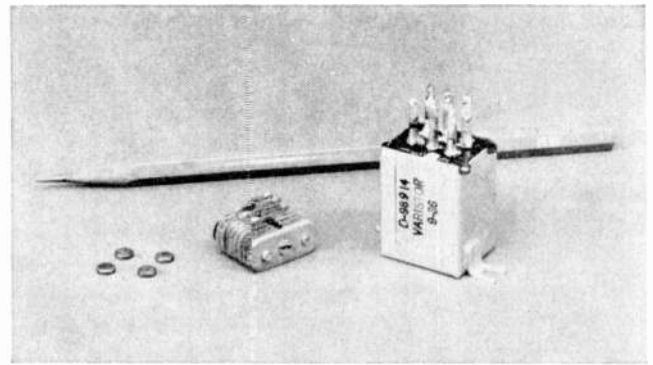
COPPER-OXIDE VARISTOR USED AS MODULATOR AND DEMODULATOR

Many types of carrier telephone systems employ copper-oxide varistors instead of vacuum tubes in their modulating and demodulating circuits. These devices are capable of accomplishing essentially the same results.

The principle of the copper-oxide varistor as a rectifier for converting alternating to direct currents in power supply circuits was discussed briefly in Article 7.5 and the underlying theory was further examined in Chapter 23. Its essential characteristic for the present purpose is that, as shown in Figure 27-6, its resistance varies with the magnitude and polarity of the applied voltage. This is a typical curve for a single disc-shaped copper-oxide unit having a diameter of $\frac{3}{16}$ inch. It will be noted that the resistance of the unit varies from a relatively low value when the copper is negative with respect to the copper oxide, to a very high value when the voltage polarity is reversed.

For use as modulators and demodulators in carrier systems, four of these tiny copper-oxide units are mounted in a sealed container having a maximum dimension of less than one inch. The characteristics of such units are very stable and their useful life is apparently indefinite under normal operation.

In the channel modulator and demodulator circuits of most carrier systems, the varistor units are connected in the Wheatstone bridge arrangement illustrated in Figure 27-7. (In the symbols used here for the varistor units, the copper-oxide is represented by the arrow, and the copper by the crossbar. The conducting direction of the unit is thus in the direction of the arrow point.) The carrier voltage, C, is made very large as compared with the signal voltage, V, so that the resistance presented by the varistor units is effectively under the control of the carrier voltage alone. In

other words, the resistance of the varistors varies from a low value to a high value at the frequency of the applied carrier voltage.

Under these circumstances, the network of varistors will act to virtually short-circuit the line during the positive halves of the carrier voltage cycle; and to present an open circuit across the line during the negative halves of the carrier voltage cycles. This is illustrated by the two diagrams of Figure 27-8 where the varistors are indicated as perfect conductors during the positive pulse and as opens during the negative pulse. The effect on the applied signal voltage, V, is therefore to block it completely during the positive half of the carrier cycle and to permit its free transmission during the negative half of the carrier cycle. The varistors thus act effectively like a switch, opening and closing at the frequency of the carrier voltage. The resultant output current is shown in Figure 27-9.

An analysis of this current curve would show that its principal components are the signal frequency and the upper and lower side-bands of the carrier frequency. If we assume for the signal voltage a sine wave of the form—

$$\epsilon = A \sin Vt$$
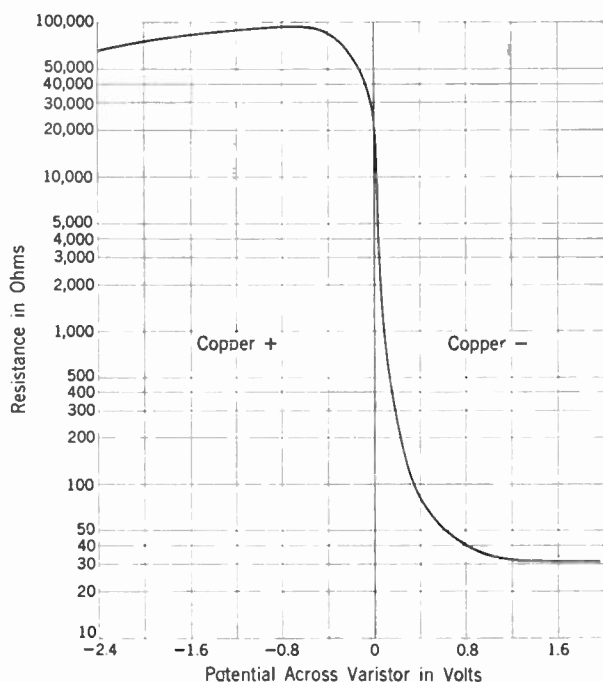
FIG. 27-6 RESISTANCE CHARACTERISTIC OF COPPER-OXIDE VARISTOR
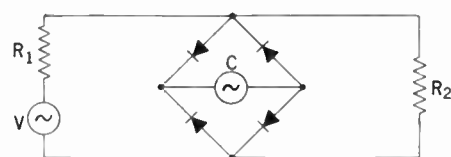
FIG. 27-7 BALANCED BRIDGE MODULATOR CIRCUIT

where A represents the amplitude of the signal and V is the signal frequency, an approximate equation for the output current represented by Figure 27-9 may be written as follows:

$$I = \frac{A \sin Vt}{2(R_1 + R_2)} + \frac{2A}{\pi(R_1 + R_2)} [\sin Vt \sin Ct$$
$$+ \tfrac{1}{3} \sin Vt \sin 3Ct + \tfrac{1}{5} \sin Vt \sin 5Ct + \ldots] \quad (27:9)$$

Here R_1 and R_2 are respectively the input and output resistances as indicated in Figure 27-7 and
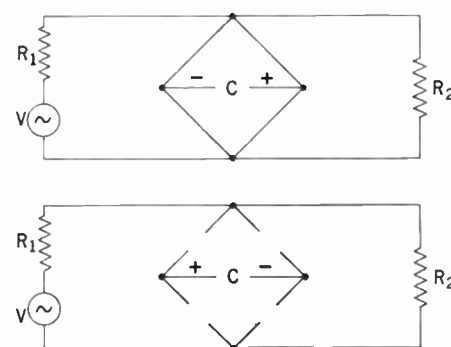
FIG. 27-8 OPERATING PRINCIPLE OF CIRCUIT OF FIG. 27-7

C is the carrier frequency.

Making use of the trigonometric relationship—

$$\sin \theta \sin \phi = \tfrac{1}{2} \cos (\theta - \phi) - \tfrac{1}{2} \cos (\theta + \phi),$$

the above equation may be rewritten as—

$$I = \frac{A \sin Vt}{2(R_1 + R_2)} + \frac{A}{\pi(R_1 + R_2)} [\cos (C - V)t$$
$$- \cos (C + V)t + \tfrac{1}{3} \cos (3C - V)t$$
$$- \tfrac{1}{3} \cos (3C + V)t + \tfrac{1}{5} \cos (5C - V)t$$
$$- \tfrac{1}{5} \cos (5C + V)t + \ldots] \quad (27:10)$$

The first term of this equation represents the original signal voltage with a reduced amplitude. The first two terms inside the brackets are the lower and upper side-bands of the modulated carrier wave, and the remaining terms in the brackets represent similar upper and lower side-bands of odd multiples of the carrier frequency. The equation does not include any term for the

[257]
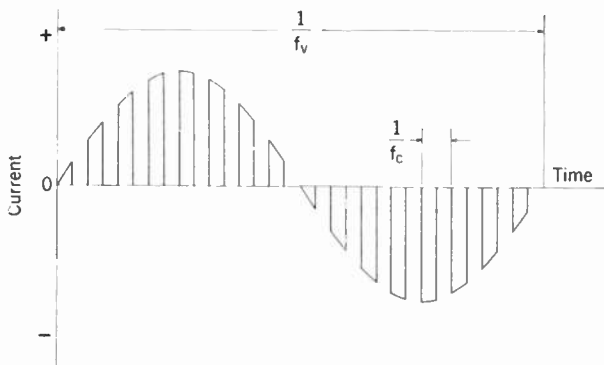
FIG. 27-9 OUTPUT CURRENT OF BALANCED BRIDGE MODULATOR

carrier frequency itself, showing that the carrier is suppressed by the balanced arrangement of the varistors.

In practice, only one of the side-bands of the carrier frequency is made use of in most cases and this is selected from the several frequency terms appearing in the output by means of a suitable band-pass filter. A demodulator arrangement, identical to that shown in Figure 27-7, is used at the receiving end of the carrier line to restore the original signal frequency. In this case, the frequencies applied to the varistor circuit (demodulator) are the received side-band and a locally generated carrier identical in frequency to that supplied to the modulator at the sending end. Thus, if we assume that the lower side-band is transmitted, the signal frequency applied to the demodulator may be indicated in the form, $K \cos (C - V)t$. When this term is substituted in equation (27:9) in place of $A \sin Vt$, the first term inside the brackets in equation (27:10) will become:

$$\cos [C - (C - V)] t = \cos Vt$$

This is the desired original signal and it can be selected from the other components of demodulation by the use of a simple low-pass filter.

For the group modulators and demodulators of
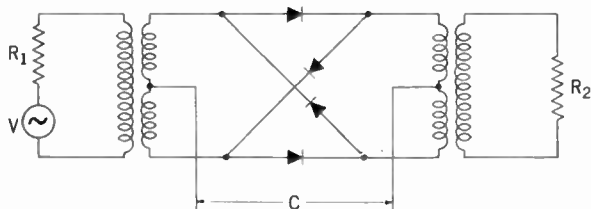
FIG. 27-10 LATTICE MODULATOR CIRCUIT

broad-band carrier systems, a somewhat different arrangement of the varistor units is frequently employed. This is illustrated in Figure 27-10. It is also a balanced bridge arrangement but the circuit connections and the configuration of the varistors are such that, as indicated in Figure 27-11, the signal voltage is impressed across the output transformer in one direction during one-half of the carrier cycle, and in the other direction during the other half of the carrier cycle. In other words the circuit acts like a reversing switch operating at the carrier frequency and results, in the ideal case, in the output current wave shown in Figure 27-12.

Using the same terminology as in the preceding discussion, the approximate equation for the curve
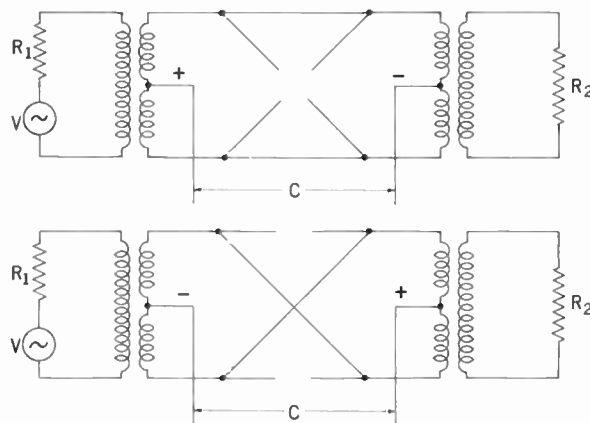
FIG. 27-11 OPERATING PRINCIPLE OF CIRCUIT OF FIG. 27-10

of Figure 27-12 is—

$$I = \frac{2A}{\pi(R_1 + R_2)} |\cos (C - V)t - \cos (C + V)t$$
$$+ \tfrac{1}{3} \cos (3C - V)t - \tfrac{1}{3} \cos (3C + V)t$$
$$+ \tfrac{1}{5} \cos (5C - V)t - \tfrac{1}{5} \cos (5C + V)t + \ldots]$$

$$(27:11)$$

Comparing this equation with (27:10), it will be noted that the desired side-bands are still present in the first two terms in the brackets, and the carrier is likewise suppressed. The signal frequency term, however, is no longer present. Moreover, the amplitudes of the side-bands are twice as great as in the previous case. This modulator therefore has the advantage of automatically suppressing the unwanted signal frequency components and of providing a larger output of the desired side-

bands. These characteristics are particularly desirable in group modulators where the wide band transmitted makes maximum side-band output, and the reduction of the number of unwanted products, very important. This arrangement of course operates as a demodulator in exactly the same way and has the same advantages.

In both of the examples of copper-oxide modulator operation discussed above, it was assumed for the sake of simplicity that the varistors acted as perfect rectifiers and were perfectly balanced in the bridge configuration. In practice, this ideal condition can only be approximated. The varistors do not actually present zero resistance to the transmission of current in one direction and infinite resistance to transmission in the other direc-
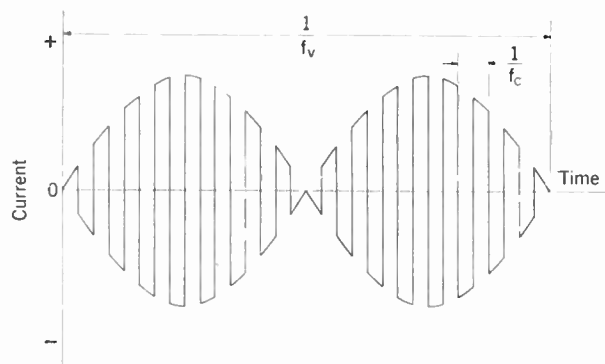
FIG. 27-12 OUTPUT CURRENT OF LATTICE MODULATOR

tion. Nor, as may be seen from Figure 27-6, is the transition from high resistance to low resistance as sharp as might be desired. Exact balance between the four varistors in the bridge connections is also a condition which can only be approached in practice.

As a result of the above practical facts, the modulator and demodulator outputs always contain numerous components additional to those indicated by equations (27:10) and (27:11), including the carrier frequency itself. Most troublesome of these unwanted components, probably, are harmonics of the signal frequency which may fall within the range of the useful side-band and thus cause distortion. Except for such frequencies as this, the unwanted components can be completely eliminated by means of suitable filters. However, it is of course desirable that as large a part as possible of the total output energy should

appear in the wanted components. This result can be effected to a considerable degree by properly proportioning the values of the applied signal and carrier voltages. Finally, it is worth noting that where greater output energy is required, each varistor can be made up of a number of individual units or discs connected in multiple or series-multiple.

Modulation in radio systems ordinarily involves substantially larger amounts of power than in the carrier systems discussed above. Considerations of power efficiency, therefore, are of much greater importance. Also, because of the much larger power requirements, vacuum tubes must always be used in the modulating circuits since any practical varistor arrangement would be inadequate to handle the required power. Power efficiency, in this case, also demands that relatively more attention be given to the percentage or extent of the modulation. This is measured by the ratio of the peak signal voltage to the peak carrier voltage, which is known as the **modulation factor**. This factor normally ranges in value from 0 to 1. At zero no modulation exists and the modulator output is simply the carrier wave. A modulation factor of 1 corresponds to 100% modulation, meaning that the modulated wave varies in amplitude at the peaks from twice the amplitude of the carrier wave to zero amplitude. In other words, the maximum signal amplitude is equal to the carrier amplitude. Greater modulation than this, a condition known as over-modulation, would of course indicate a modulation factor greater than 1 and would result in distortion because the modulated wave would not be continuous.

The total power of an amplitude-modulated carrier wave is equal to that of the carrier plus that of the two side-bands. It can be shown that the power in the side-bands is proportional to the square of the modulation factor, and with 100% modulation, the total power of the modulated wave is 50% greater than that of the carrier wave; i.e., one third of the total output power is contained in the side-bands. With lesser modulation, the side-band power drops off rapidly. Since it is only the power in the side-bands that carries the signal, it is clear that for efficient transmission, the modulation ratio should be held as close to 100% as possible.

The general type of vacuum tube modulator that was discussed earlier in connection with carrier systems, in which the carrier and signal wave were applied to the grid circuit, is sometimes used

in relatively low power radio transmitters. In radio terminology, such a circuit would usually be known as a grid-modulated Class C amplifier. More generally used in radio work is the plate-modulated Class C amplifier illustrated in Figure 27-13. In this circuit, the carrier voltage is applied to the grid as shown. When no signal voltage is applied, the plate current consists of a series of positive pulses, as indicated in (A) of Figure 27-14, because of the large negative grid bias. When a signal voltage is applied to the plate circuit in series with the B battery as shown, the total plate voltage varies above and below the fixed B battery voltage in accordance with the signal. This is illustrated in (B) of Figure 27-14. The varying plate voltage causes the plate current to vary as indicated in (C) of Figure 27-14, where the envelope of the wave has the form of the signal. The current pulses produce an oscillatory current flow in the tank circuit, which results in the output voltage wave shown at (D).
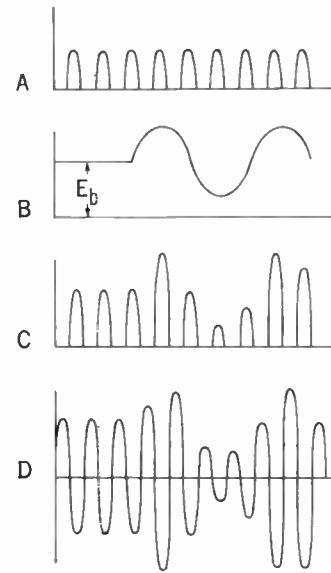
FIGURE 27-14
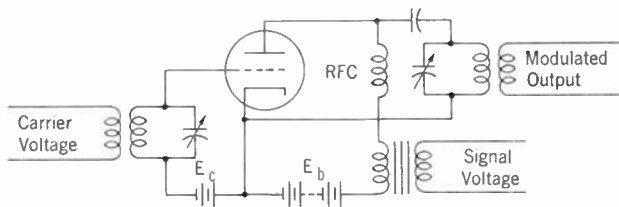
FIG. 27-13 PLATE-MODULATED AMPLIFIER

Examination of Figure 27-14 will show that for 100% modulation (modulation factor of 1), the maximum amplitude of the applied signal voltage must be equal to the value of the d-c plate supply voltage. This means that 50% of the total r-f output power must be furnished by the signal source at 100% modulation. The requirement that this large portion of the power must come from the modulating source has certain disadvantages in comparison with a grid-modulated transmitter in which all the output power is furnished by the d-c plate-supply.

Demodulation in radio systems involves the same principles as were outlined above in connection with carrier systems. But the particular devices used in this case are generally known as detectors. Such devices were discussed in Chapter 26.

27.3 Frequency Modulation

The use of frequency modulation is confined en-

tirely to radio systems operating in the very high frequency band or above, where it has certain definite advantages over amplitude modulation in minimizing interference from "static" and extraneous signals. It depends upon varying the frequency of a carrier wave of fixed amplitude above and below a central or normal frequency in accordance with the amplitude variations of an applied signal voltage. The process is roughly illustrated by the three wave diagrams of Figure 27-15. The amount of frequency change that is produced by the signal is called the **frequency deviation** and, ideally, this should be as high as possible in order to obtain the maximum signal to noise ratio. However, since it is obvious that the total bandwidth of the modulated wave to be transmitted will increase with increases in the maximum frequency deviations on both sides of the unmodulated carrier frequency, it is necessary as a practical matter to arbitrarily limit the maximum permissible deviations to values that will keep the total bandwidth that must be assigned in the radio spectrum to each FM channel as narrow as feasible. The maximum permissible deviation has been specified by the Federal Communications Commission at 75 kc for FM broadcasting, and at 15 kc for such applications as mobile radio service.

As in amplitude modulation, frequency modulation results in a modulated wave containing the carrier frequency and other frequencies above
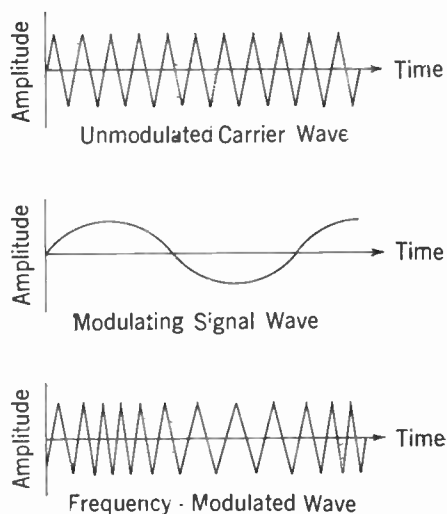
FIGURE 27-15

and below the carrier frequency. If, for purposes of analysis, we assume a sinusoidal signal wave of frequency f_s impressed on a carrier of frequency f_c, it may be shown that the instantaneous value of the modulated resultant wave may be written as—

$$e = E_m \sin\left(\omega_c t + \frac{\Delta f}{f_s} \sin \omega_s t\right) \qquad (27:12)$$

where E_m represents the constant voltage amplitude of the carrier wave, Δf the frequency deviation, and ω_c and ω_s are $2\pi f_c$ and $2\pi f_s$ respectively. This equation can be expanded into the following infinite series—

$$e = E_m\{A_0 \sin \omega_c t + A_1 [\sin (\omega_c + \omega_s) t$$
$$- \sin (\omega_c - \omega_s) t] + A_2 [\sin (\omega_c + 2\omega_s) t$$
$$+ \sin (\omega_c - 2\omega_s) t] + \ldots\}$$

$$(27:13)$$

This equation shows that in addition to the carrier frequency itself, the modulated wave includes an infinite series of side frequencies having values equal to the carrier frequency plus and minus the signal frequency and all of its integral multiples. The relative magnitude of the carrier frequency and the side frequencies depends upon the ratio of the frequency deviation to the signal frequency. This ratio is commonly designated m_f, and is known as the **modulation index**.

In general, the number of side frequency pairs that must be transmitted to carry a signal with reasonable fidelity increases with the value of the modulation index. While the number of side-pairs

is theoretically infinite, only side frequencies containing 1% or more of the total energy are considered significant in practice. This limits the total number of frequencies necessary for satisfactory transmission within reasonable bounds. For example, Figure 27-16 shows the relative amplitudes and positions in the spectrum of a representative FM wave where an 80 mc carrier is modulated with a 4 kc signal when the maximum frequency deviation is 40 kc. It will be noted that the amplitudes of the carrier and side frequencies vary according to an irregular pattern, but tend to decrease in value rapidly at the extremities. The values of these amplitudes of course correspond to the constants A_0, A_1, A_2 . . . of equation (27:13). Their calculation involves the use of certain complex mathematical functions, but tables are to be found in the handbooks showing their values for various values of modulation index, m_f.

It will be observed that the frequency deviation in this example is only 40 kc, although a maximum deviation of 75 kc is permitted. This is the more usual situation because the extent of deviation is determined by the signal amplitude and will reach maximum values only at peak signal strength. The total bandwidth required by the frequency-modulated carrier at any time thus depends both upon the intensity of the applied modulating signal and its frequency, since m_f is equal to $\Delta f/f_s$. In the example, the total swing is shown as 112 kc, which, it will be seen, is appreciably more than twice the deviation of 40 kc. At higher values of m_f, however, the total required bandwidth approaches closer and closer to $2\Delta f$.

At lower frequencies in the voice range, the modulation index will be high and the number of

$$f_c = 80 \text{ MC}$$
$$f_s = 4 \text{ KC}$$
$$\Delta f = 40 \text{ KC}$$
$$m_f = \frac{\Delta f}{f_s} = 10$$
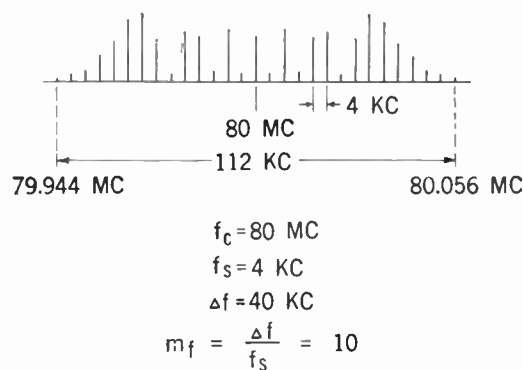
FIG. 27-16 INSTANTANEOUS FREQUENCY COMPONENTS OF F-M WAVE

[261]

significant side frequencies will be corresponding-
ly large. But the separation of these fre-
quencies in the spectrum is equal to the signal
frequency so that the required bandwidth is
generally no greater than 150 kc, twice the maxi-
mum permissible deviation. At an extreme high
voice-frequency—say 15 kc—a maximum-inten-
sity signal would result in a modulation index of
75/15 or 5. In this case, the theoretical required
bandwidth would be 240 kc. This is an extremely
unlikely situation, however, because the amplitude
of any normal high-frequency signal is substan-
tially less than that of the low-frequency compo-
nents, and maximum frequency deviation at 15 kc
would not be expected to occur. Practice has
shown that a total bandwidth of 200 kc is suffi-
cient for high fidelity transmission of audio sig-
nals, and this is the bandwidth designated by
the F.C.C. for FM radio broadcasting channels.
Where the permissible deviation is less than 75
kc, as in mobile radio service, the permissible
channel bandwidth is of course reduced propor-
tionately.

Various circuits have been devised for modulat-
ing the frequency of a carrier wave in accordance
with the amplitude variations of a signal voltage.
One basic arrangement for accomplishing this
purpose makes use of what is known as a "react-
ance tube" circuit. The principle is illustrated
schematically in Figure 27-17. The oscillator cir-
cuit at the right is of the Hartley type, and its
output frequency, when the circuit is open at
points 1 and 2, is determined by the constants L_1
and C_1 of the tuned tank circuit. When connected
to the reactance tube circuit, the oscillator circuit
presents a purely resistive load to the output of
the reactance tube circuit at the operating fre-

quency. The reactance tube is bridged across
points 1 and 2 as is also the voltage divider, C-R.
The total impedance of C-R is high, and the value
of C is so selected that its reactance in the car-
rier frequency range is much higher than the re-
sistance of R. The oscillator sets up a small r-f
current through C-R, which, because of the domi-
nating capacitive reactance, will lead the voltage
across 1 and 2 by almost 90°. The voltage across
R, which is applied to the grid of the reactance
tube, will also lead $E_{1,2}$ by the same angle. Since
the plate current of a vacuum tube is in phase
with the grid voltage, the output current of the
reactance tube must also lead $E_{1,2}$ by 90°. The re-
actance tube circuit accordingly acts effectively
like a pure capacitive reactance connected across
points 1,2. Since this reactance is in parallel with
the capacitance C_1 of the oscillator tank circuit,
it will affect the frequency of the oscillator. The
magnitude of the bridged capacitive reactance is
determined by the value of the plate current in
the reactance tube, which is in turn controlled by
the value of the voltage applied to its grid. When
the signal voltage is connected across R as shown,
therefore, the net reactive capacitance of the re-
actance tube circuit will vary as the signal voltage
amplitude varies, and the frequency of the oscil-
lator will vary accordingly.

Because it is controlled by simple inductive and
capacitive elements, the frequency of the oscilla-
tor of Figure 27-17 would ordinarily not be suf-
ficiently stable to meet the exacting requirements
of radio transmission. This difficulty is overcome
in one type of FM transmitter by means of an
automatic frequency control circuit employing a
discriminator and a fixed crystal-controlled auxil-
iary oscillator. In the arrangement most widely
used in FM broadcasting transmitters, the basic
oscillator frequency is controlled by a crystal.
This makes it impossible for the signal voltage to
cause the oscillator frequency to vary. It can,
however, cause variations in the phase angle of
the oscillator output corresponding to the ampli-
tude variations of the signal. By the use of appro-
priate phase-shifting networks, the "phase-modu-
lated" output is then translated into an equivalent
frequency modulation.

In practice, the basic oscillator is operated at
a much lower frequency than the actual carrier
frequency radiated from the transmitting an-
tenna, but is brought up to the desired value by
several stages of frequency multiplication. The
multiplication process naturally increases the de-
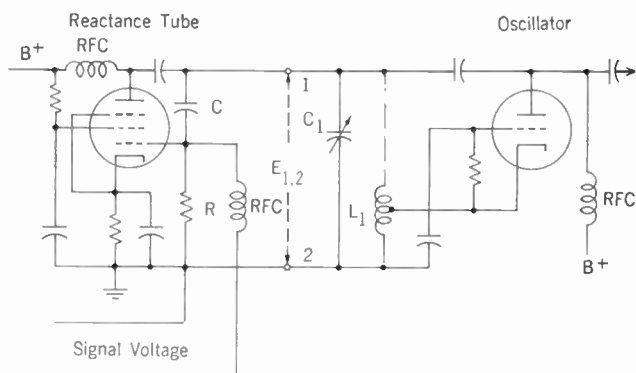
FIG. 27-17 FREQUENCY MODULATION WITH REACTANCE
TUBE CIRCUIT

viation frequencies in the same way that it increases the mean carrier frequency.

27.4 Pulse Modulation

Although applications have been limited in telephone practice up to this time, it is possible to transmit telephone signals by means of a series of pulses of energy. The conversion of the telephone signal into energy pulses from which it can be reproduced at the receiving end is known as **pulse modulation**. The method depends on the fact that any signal waveshape can be completely defined by taking very short (essentially instantaneous) samples of its amplitude at regularly spaced successive time intervals, such that at least two samplings are taken for each cycle of the highest frequency included in the signal wave. For example, if a voice-wave has been passed

Modulating Signal

Pulse - Amplitude Modulation

Pulse - Width Modulation

Pulse - Position Modulation
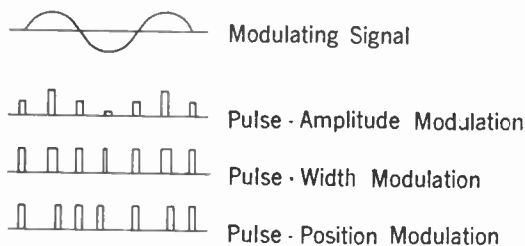
FIG. 27-18 PULSE MODULATION METHODS

through a low-pass filter having a cutoff frequency of 4000 cycles, all of the information required for its reproduction will be contained in a regular series of samples taken at the rate of 8000 per second.

The transmission of information by means of discrete energy pulses has certain definite advantages in radio systems particularly. It permits the utilization of amplitude limiting techniques by means of which noise and other interference, including fading, may be minimized. It also tends to reduce power consumption at the radio transmitter because the energy output is not continuous. Moreover, it makes possible the transmission of a number of separate signals over a single carrier by means of **time-division multiplex**. That is, the intervals between the successive pulses of a given signal can be employed to transmit comparable sampling pulses of other signals.

There are several possible ways in which the signal sampling information can be incorporated in the energy pulses. Three of these are illus-
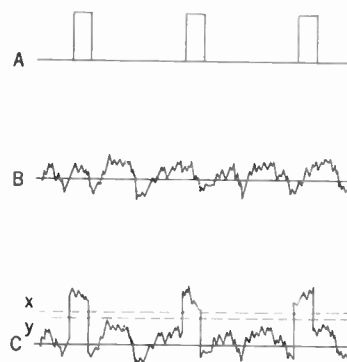
FIGURE 27-19

trated in Figure 27-18. In pulse-amplitude modulation, the sampling pulses are of uniform time duration, but vary in amplitude in accordance with the signal amplitude at the sampling instants. In pulse-width modulation and pulse-position modulation, the pulses are of constant amplitude but vary with the signal amplitude in duration or width in the first case, and in position with respect to a uniform time scale in the second case.

Because of the constant amplitude, the latter two types of pulse modulation have the advantage that limiting circuits can be used at the receiving end to practically eliminate the influence of line noise. This is illustrated by Figure 27-19 where C represents a received wave consisting of the sum of the original pulse wave A and the noise B which may have been superimposed along the transmission line. By use of methods discussed in Article 26.5, both top and bottom of the irregular wave of C can be clipped to leave only the undisturbed sections between lines x and y. These clean sections may then be amplified, if desired, to create an exact reproduction of the original pulse signal A. Thus, so long as the pulse signal is appreciably stronger than the maximum line noise, the latter is completely removed.

It may be evident that transmission of information by pulse systems has much in common with ordinary telegraph systems, except that the repetition rate of the pulses is very much higher. When modulated pulses of constant amplitude are superimposed on a high frequency carrier by ordinary amplitude methods, the transmitted wave consists merely of a series of spurts of the carrier frequency just as in simple carrier telegraph systems. Transmission of amplitude-modulated pulses would of course involve variations in the amplitudes of the carrier frequency spurts. This

Actual Signal Wave

Quantized Wave

Amplitude

Time
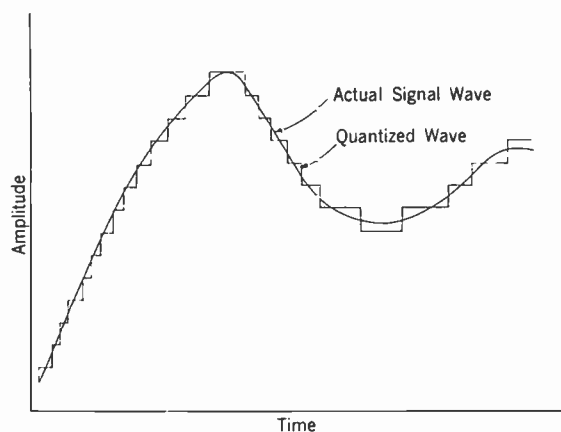
FIGURE 27-20

technique may nevertheless have its advantages by permitting the use of simpler and less costly transmitting and receiving apparatus in certain cases.

Both pulse-width and pulse-position systems impose rather severe requirements on the receiver in that their accurate detection demands exact measurement of very small time intervals. There is an obvious advantage, if possible, in employing pulses of uniform amplitude which the receiver is only required to recognize as being there or not being there—simply "yes or no". This advantage is attained in the system known as **pulse-code modulation.** Here each item of information that was represented by a single pulse in the systems discussed above is coded into a series of "off or on" pulses similar to the codes used to represent letters in teletypewriter systems. The amplitude of the basic signal being sampled, however, may vary continuously over a wide range and may thus have an infinite number of values. To keep the total number of codes within reasonable limits, it is necessary to divide the total amplitude range of the basic signal into a limited number of finite steps or "quanta" as indicated by Figure 27-20, with a separate code assigned to each quantum step. Each code then represents any signal amplitude in a range between two successive steps.

The quantizing process necessarily produces some distortion or "granularity" in the coded signal sent over the line. It also produces some "quantizing noise" at the receiving end. It has been found in practice, however, that when as many as 128 quantum steps are employed, speech signals are reproduced with a high degree of fidelity. In a binary or off-or-on system, 128 separate codes require the use of seven positions or "bits" per code as distinguished from the five positions of teletypewriter operation where only 32 different codes are necessary (Figure 11-4).

At the receiving end, each 7-element code signal is translated into the single amplitude pulse which the code represents. The successive amplitude pulses are then applied to a low-pass filter (cutting off at 4000 cycles in the case of a voice wave) the output of which will be an exact copy of the original signal wave except for the quantizing effect. Successful operation of course requires the continuous maintenance of precise time synchronization between receiver and transmitter. This is accomplished in the same general manner as in teletypewriter systems by transmitting frequent synchronizing pulses which are additional to the message-carrying code pulses.

Since the individual transmitted pulses are necessarily of extremely short duration (microseconds or fractions of micro-seconds), it is obvious that all operations must be performed by purely electronic rather than electro-mechanical means. It is also fairly evident that transmission must involve a very wide band of frequencies. The shorter the individual pulse, the greater the required bandwidth. A system arranged for handling twenty-four telephone channels by time division multiplex, for example, must transmit at least 1,536,000 bits or items of information each second—8000 (signal samples) times 24 (multiplex channels) times 8 (7 code elements plus a synchronizing pulse). The individual pulses in such a case would obviously have to be appreciably less than one micro-second in duration.

To counteract the high losses associated with such wide band transmission over ordinary types of facilities, it is necessary to employ repeaters at frequent intervals along the line. Because of the simplicity of the signal (pulse or no pulse), however, the repeaters may be of a simple regenerative type, operating on the same principle as the regenerative repeaters of telegraph practice. And like the telegraph repeater, they have the great virtue of delivering a perfect output signal even when the input signal is accompanied by considerable noise or other distortion. There is therefore no cumulative noise effect and pulse code signals can thus be repeated any number of times and still deliver message signals at the final receiving point which are completely free from noise or other distortion, except for the slight quantizing noise previously mentioned.

[264]

CHAPTER 28

LONG DISTANCE TRANSMISSION SYSTEMS
VOICE-FREQUENCY TELEPHONE CIRCUITS

28.1 Types of Telephone Transmission Systems

The various kinds of line facilities and apparatus that we have considered in preceding Chapters are applied in practice to the development of several distinct types of long distance telephone circuits. Such circuits may be broadly classified as between those which operate at voice frequencies and those which operate at higher—carrier or radio—frequencies. In the former group are the ordinary 2-wire telephone circuits which employ a single pair of open wire or cable conductors as the transmitting medium, as is the general practice in the case of most local and short-haul toll telephone service. The voice-frequency group also includes the 4-wire cable circuits in which a separate pair of cable conductors is used for transmission in each direction. If they are of considerable length, both the 2-wire and 4-wire circuits require the insertion of telephone repeaters at regular intervals in order to maintain transmission at satisfactory levels.

Except in the case of coaxial conductors, carrier circuits employ the same or similar kinds of wire facilities for transmission as do voice-frequency circuits. They also require the use of amplifiers or repeaters at regular intervals along the line. This latter statement also applies to superhigh-frequency overland radio circuits where relay stations are necessary at regularly spaced intervals along the route.

It is not possible to make an unqualified statement as to the particular kinds of situations in which each of the above types of circuits may be best applied in practice. In general, however, 2-wire circuits are commonly used for relatively short distances—not more than a few hundred miles maximum. Four-wire cable circuits are used for somewhat longer distances when equipped with "echo suppressors". Carrier or radio circuits are generally used for the longest distances, although their use is not limited to such application.

In this and following Chapters, some of the more essential problems involved in the design and operation of long voice-frequency, carrier, and radio circuits are considered.

28.2 Repeater Spacing

The different types of circuit facilities, whose principal characteristics are discussed in Chapter 18, show a wide diversity in their relative transmission efficiencies. It may be noted, for example, that at voice frequencies a loss greater than one db is caused by one mile of 19-gage non-loaded cable side circuit while a 165 open wire phantom circuit causes a loss of only .025 db per mile. In other words, one mile of the former gives rise to as great a loss as nearly forty miles of the latter.

Prior to the advent of the telephone repeater in 1915, large gage open wire facilities were used for all very long circuits; furthermore, such facilities were usually loaded. But even with the use of loaded 165 facilities, the maximum practicable range for long distance telephony was limited to about two-thousand miles. The application of the telephone repeater had two fundamental and far reaching effects—first, it made possible an indefinite extension of the maximum range of telephonic communication; and second, it permitted smaller wire gages for long distance service and so helped to make economically and physically feasible the great expansion in the number of long distance circuits that has occurred since its introduction.

The first transcontinental telephone service was furnished by loaded 165 open wire facilities with repeaters inserted at 500 to 600-mile intervals. This same service was later improved by removing the loading from the open wire facilities and reducing the repeater spacing. Here the repeater served another purpose; it improved the quality of the circuit by making possible the elimination of the inherently troublesome open wire loading. Repeaters are now used in practically all long distance cable and open wire circuits. Since open wire facilities must for mechanical reasons be of relatively large gage and suspended with considerable separation between conductors, their resistance and capacitance values are relatively low. As a result repeaters need only be spaced at intervals of the order of 150 to 350 miles to compensate for the energy attenuation caused by the conductors. This means that even in the longest

[265]

circuits the number of repeaters in tandem is not very great. On the other hand, in cable facilities the wire conductors are usually of 16 or 19-gage and even though loading is used, repeaters must be inserted at 50 to 60-mile intervals. It follows that a very long cable circuit must include a considerable number of repeaters in tandem. In either case, it is the usual practice to employ repeaters at the terminals as well as at intermediate points along the circuit.

In an open wire circuit, 165 wire will generally have the best electrical and mechanical characteristics, with 128 wire circuits next, and 104 wire circuits last. By increasing the repeater spacing when larger wires are used, these three types of circuits can be made practically identical from a transmission standpoint. However, most open wire lines include different sizes of wire and the spacing of the repeaters is determined by the losses of the smallest wires. Nevertheless, even though short repeater sections are used, climatic conditions may be such as to put occasional severe strains on the wires and thus necessitate the use of 128 wire or even 165 wire to obtain greater mechanical strength than is possible with 104 wire. The final decision as to the size of the open wire, and hence the repeater spacing, must be based on the proper consideration of both the economic and electrical factors. The latter include repeater balance, transmission variation due to temperature and other weather changes, and echo effects, all of which are discussed in following Articles. In practice it has usually been found that the repeater spacing on open wire facilities should not exceed 350 miles for 165 wire, 225 miles for 128 wire, and 150 miles for 104 wire.

Long toll cables usually employ 19 and 16-gage conductors for voice-frequency transmission. The former gage is used for both 2 and 4-wire circuits, while the latter is frequently used for program services (radio broadcasting networks) and to some extent for 2-wire message circuits. These conductors are loaded to reduce their attenuation and thereby permit longer repeater spacing. The type of loading used depends upon the lengths of the circuits and the uses that are made of them. Although this might imply that it is desirable to have a different type of facility for each length and circuit use, it has been found practicable to obtain satisfactory results with only a few standard types. The characteristics of all of these are such that the preferred repeater spacing is about 50 miles for aerial cable and 55 miles for under-

ground and buried cable.

These requirements, however, are not so rigid as to preclude a needed element of flexibility. Thus an open wire or cable route obviously will not have towns and cities located exactly at the points where it may appear desirable to locate the repeaters. Within limits, the repeater spacing may be varied somewhat to conform with the preferable location of the repeater stations.

28.3 Repeater Gains and Transmission Levels

After the location of the individual repeater stations has been selected, the amount of gain to be inserted in each circuit at each repeater point must be determined. It is generally desirable to keep the energy of the message currents at the highest practicable level in order to minimize noise interference. If the transmission level of the message currents is too low, any small noise currents that may be induced into the circuit from external sources may be relatively great enough to cause excessive interference when they are amplified by the repeaters along with the message currents. This can be overcome by keeping the transmission level of the message currents high with respect to the level of the induced noise currents. However, it should not be forgotten that there is a limit to the amount of energy that any particular amplifying circuit can handle, and even before reaching this limit, distortion is introduced.

In adjusting the gains of 22-type repeaters, the ordinary limits are as follows: With the volume of transmission at the switchboard at the sending terminal of a circuit defined as "zero transmission level", a 22-type repeater may ordinarily be operated to deliver a volume of transmission not exceeding the zero level by more than 6 db, and with a maximum net gain of not more than 18 db. The output of V-type repeaters presented to the input of the hybrid coil (which in this case is not a part of the repeater) may be as high as 10 db above zero level and the total gain of the repeater may be permitted to approach 25 db.

In repeaters used in 4-wire cable circuits, much higher gains are possible. However, crosstalk considerations usually prevent the attainment in practice of the maximum possible gain. It is permissible to operate these repeaters between an input volume not lower than about 25 db below zero level and a delivered output not greater than about 10 db above zero level. This means a possible gain

of 35 db, under which condition the energy de-
livered is nearly two-thousand times as great as
the energy received. This extreme energy ratio
is the reason for the crosstalk limitation. If an
incoming cable pair is adjacent to an outgoing
cable pair of another 4-wire circuit and there
exists a small crosstalk unbalance from one pair
to the other, the highly energized circuit may
transfer a quantity of energy which although an
almost negligible fraction of its own energy, may
nevertheless be quite appreciable as compared
with the energy in the other circuit, which is only
about 1/2000th as great in value. This crosstalk
energy is applied to the repeater with the incom-
ing transmission and is amplified along with and
to the same degree as the incoming transmission,
thereby tending to become audible.

In laying out long circuits containing a number
of repeaters in tandem, use is made of an "energy
level diagram" which shows in a single chart not
only the losses in each line section and the gain
of each repeater, but also the level of the voice
energy at each point along the circuit as com-
pared with the energy originally applied to the
circuit terminal (zero level). Figure 28-1 gives
such a diagram for a typical 2-wire cable circuit.
The ordinates represent energy levels in decibels
above and below zero level, losses being measured
downwards and gains upwards. The gains of the
repeaters are naturally represented by straight

vertical lines, while the line and equipment at-
tenuation losses are indicated by lines between re-
peater stations sloping downward in the direction
of transmission. A separate set of zigzag lines is
required to show transmission in each direction,
even when the net equivalent of the circuit and
the gains of each repeater are the same in each
direction. Such a chart is valuable not only in
facilitating the original engineering design of the
circuit, but also as maintenance information to
enable the repeater stations to know both the gain
at which each repeater is to be operated and the
proper output energy level for transmission in
each direction.

It will be noted that in the case of the 22-type
repeaters, a single vertical line is used to repre-
sent the net gain, which is made up of the dif-
ference between the gain of the amplifier itself
and the losses of the hybrid coils that are part of
the repeater. In the case of the V-type repeater,
on the other hand, the hybrid coil losses are indi-
cated separately from the amplifier gain. This is
desirable as a practical matter because the ampli-
fiers of the V-type repeaters and their gain adjust-
ing controls are separated physically from the
hybrid coils.

28.4 Return Loss and Singing Points

In 2-wire circuits, repeater gains are usually
limited by the degree of balance
which it is possible to secure
between each line and its bal-
ancing network, rather than by
the maximum energy output of
the amplifying tubes. In other
words, the allowable amplifica-
tion of a 2-wire telephone re-
peater depends upon the gains
that make the repeater circuit
oscillate or "sing", or apprecia-
bly impair quality because of
unbalance between the line and
the associated network. Def-
inite impairment of quality is
quite noticeable just before the
"singing" point is reached.

As pointed out in Article
14.7, if identical impedances
are connected to the line and
network terminals of the hy-
brid coil, no power can pass
from the series winding to the
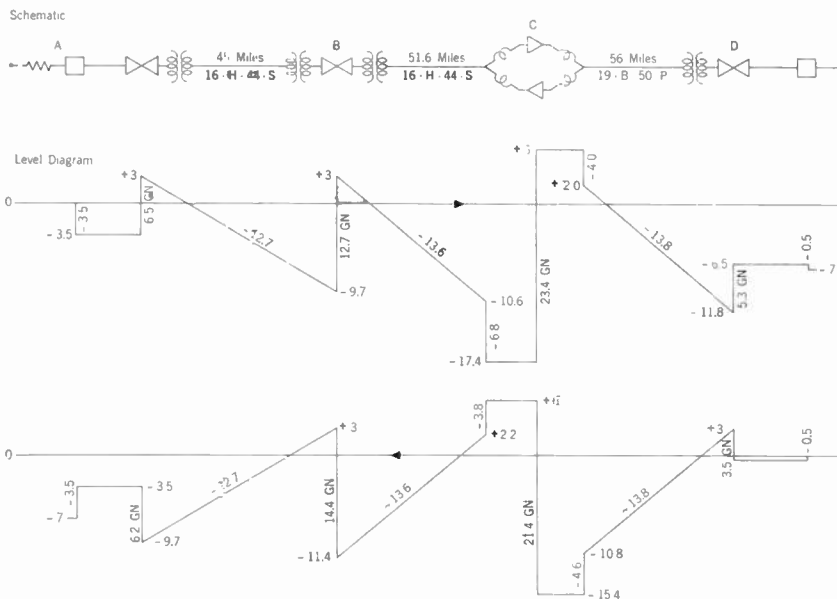
FIG. 28-1 ENERGY LEVEL DIAGRAM FOR TWO-WIRE CABLE CIRCUIT

[267]

bridge taps; in other words, there is infinite loss across this path. If, however, there is an inequality between the line and network impedances, power can pass and a finite loss may be measured between these points. The total loss across the hybrid between the series winding and the bridge taps, which may be designated L, is made up of the normal hybrid losses caused by the division of energy between the balanced circuits and a loss that is dependent upon the degree of balance between the line and network which is known as the **return loss** and designated $R.L.$ When the line impedance differs from the network impedance, a portion of the power on the line is reflected back towards the hybrid coil. The part so reflected back is less than the power sent out on the line by the amount of the return loss. The greater the departure of the line impedance from its normal value (which the network simulates), the more the power reflected back, and the smaller the return loss. This reflected power enters the hybrid coil in the same manner as normal incoming transmission and in the same way divides between the bridge taps and the series winding, thus incurring another hybrid loss. The total trans-hybrid loss then equals the $R.L.$ plus the normal loss from the series winding to the line, plus the normal hybrid loss from the line to the bridge taps. In the ordinary 22-type repeater circuit, the normal hybrid losses are about 3.25 db making the total trans-hybrid loss L equal to $R.L. + 3.25 + 3.25$. The return loss at a given frequency is thus the measured transmission loss across the hybrid coil at that frequency less the fixed losses due to the characteristics of the coil itself.

The value of the return loss is a measure of the similarity between the line and network impedances, and is the kind of quantity "singing point tests" are designed to measure to a certain approximation. Its value in db may be determined by the equation—

$$R.L. = 20 \log_{10} \frac{Z_N + Z_L}{Z_N - Z_L} \qquad (28:1)$$

where Z_N is the impedance of the network and Z_L is the impedance of the line. If the network perfectly balances the line, that is, if $Z_N = Z_L$, equation (28:1) shows that the return loss is infinite. When an unbalance exists, the loss takes a finite value. Assuming as an example that $Z_N = 600$ ohms, and $Z_L = 400$ ohms, then—

$$R.L. = 20 \log \frac{600 + 400}{600 - 400}$$

$$= 20 \log 5 = 20 \times .7 = 14 \text{ db}$$

A return loss measurement or computation is made in terms of a single frequency, and the gains and losses in the measuring circuit must be expressed for the particular frequency used since both Z_N and Z_L may vary with frequency. Such measurements or computations must be made for a number of frequencies in the voice range if it is desired to determine at what point in the range balance conditions are poorest.

In singing point tests, however, the repeater automatically selects the frequency at which singing is most likely to occur, and thus in a single measurement gives the approximate balance condition at the worst frequency. The gains of the repeater used in making the tests are ordinarily measured at 1000 cycles and for convenience these 1000-cycle values are used in determining the numerical value of the singing points. The value so determined may therefore be somewhat different from the return loss at the singing frequency because the repeater gains at this frequency may be different from the 1000-cycle gains. Phase relationships, also, may sometimes be such as to prevent singing from occurring at the frequency where the balance conditions are poorest. However, singing point tests give results sufficiently accurate for practical maintenance purposes. They provide a ready means of ascertaining what is the maximum safe working gain of a repeater when connected to a given 2-wire circuit. The measurement also gives a direct check on the effectiveness of the network balance, since a high singing point means that at no single frequency within the voice range is there an appreciable dissimilarity between the impedance of the network circuit and the impedance of the line.

Such a satisfactory balance between a line and its network depends, among other things, upon the termination of the line at the next adjacent repeater point. When making tests this termination may consist of a network or of a "passive repeater"—that is, a repeater suitably terminated in resistors so as to present its nominal impedance to the line. The balance measured under this condition is called the **passive singing point**, which means fundamentally that the test repeater is the only repeater in the circuit that amplifies the reflected power, or that only one amplification path is involved.

Now suppose that instead of being terminated at the adjacent office in a network or in a passive repeater, the circuit at that office goes through an active repeater (one in operating condition) and on to another repeater section beyond. The reflected power in the first section will still return to the test repeater, but in addition, part of the sent power will enter the second repeater, be amplified and sent into the second repeater section; if the second section contains irregularities, part of the power entering this section will be reflected back to the second repeater, through this repeater and into the first section, and then back to the hybrid coil of the test repeater, thus adding to the power returned from the first section. There are now two points in the circuit where reflected power is amplified. In other words, there are two amplification paths and with more repeaters in the circuit, there may be a third and fourth path, etc. All of these returned powers combine at the hybrid coil of the test repeater to enter the circulating path of this repeater. The greater this total power, the less the gain required to sustain singing and, accordingly, the lower the singing point. The balance for this condition is termed the **active singing point.**

28.5 Balancing Networks

From the preceding discussion it will be evident that the satisfactory operation of 2-wire repeatered circuits depends upon the impedance of the line as well as its attenuation. The extent to which the repeater may improve transmission depends directly upon the degree to which the network balances the line. In turn the degree of balance depends first upon the "smoothness" of the telephone line's impedance throughout the working range of voice frequencies, and second, upon the adjustments that it is practicable to make for the effect that terminating conditions have upon this impedance.

The basic requirements as to balance may be understood by referring to Figure 28-2. Here we have the R and X components of the characteristic impedance, Z_0, of a 104 open wire side circuit plotted (solid line) with respect to the voice-frequency band. It will be seen that the resistance component of the characteristic impedance becomes appreciably lower at the higher frequencies and that there is likewise a marked change in the value of the negative reactance.

To balance such a circuit, a network must be
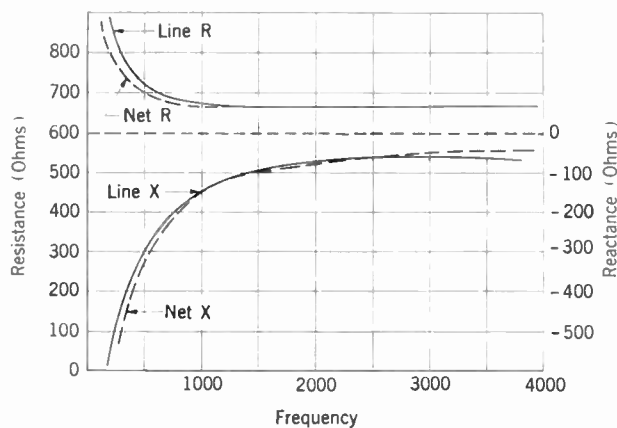
FIG. 28-2 IMPEDANCE CHARACTERISTICS OF OPEN WIRE CIRCUIT AND ITS BALANCING NETWORK

designed with impedance components that not only equal those of the line at some one frequency, but vary similarly with the impedance of the line at all frequencies within the voice band. The dashed curves in this same Figure compare the R and X components of the impedance of a network used to balance this type of line.

The essential elements of a basic network designed to balance an open wire circuit are illustrated by Figure 28-4 (A). This simple arrangement, with proper values of resistance and capacitance, will closely approximate the impedance components of the line itself. It will not, however, take care of near-end terminating conditions such as toll entrance cable, etc. Furthermore, it balances only the characteristic impedance of the circuit, i.e., the circuit must be in effect infinite in length; or in other words, terminated at the distant end in an impedance equal to the characteristic impedance. Consequently, balance, even in the open wire circuit case, involves considerations other than the mere design of a basic network that has an impedance approximating that of the characteristic impedance of the line. These balance requirements, however, are general and will be discussed after considering the basic network for the loaded cable circuit.

A basic network for a loaded circuit usually has a more complex design than a basic network for a non-loaded circuit. In this design some assumption must be made regarding the loaded circuit's near-end termination, i.e., the basic network must be chosen to balance a loaded circuit terminating at a mid-section point, or at some fraction of the loading section other than mid-section. Figure 28-3 (A) shows the resistance components of the impedance of an ideal loaded line for various

[269]

RESISTANCE

$X = .5$

$.4 \ \& \ .6$

$.3 \ \& \ .7$

$.2 \ \& \ .8$

$.1 \ \& \ .9$

$X = 0, 1.0 \ \& \ 1/2$ Coil

$$Z = \sqrt{\frac{L_0}{C_0}} \ (D \pm jQ)$$

D

f/f_c

A

REACTANCE

$$Z = \sqrt{\frac{L_0}{C_0}} \ (D \pm jQ)$$

Note: For Values of X:
From 0 to .5, Q is Positive.
From .5 to 1., Q is Negative.

Q

$X = 0 \ \& \ 1.0$

$.1 \ \& \ .9$

$.2 \ \& \ .8$

$.3 \ \& \ .7$

$.4 \ \& \ .6$

$X = .5 \ \& \ 1/2$ Coil

f/f_c

B
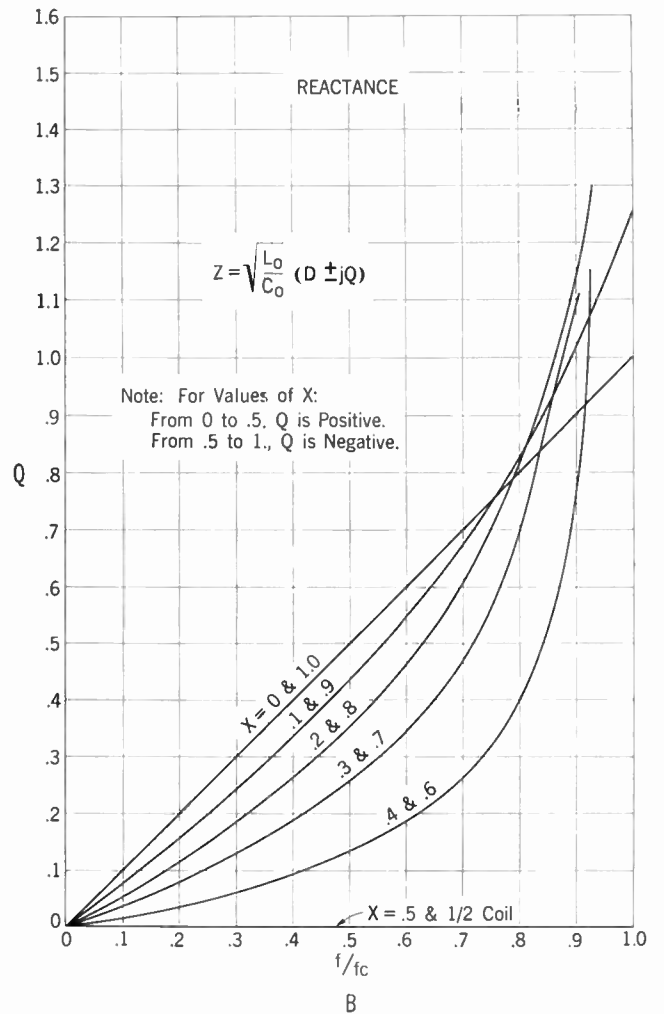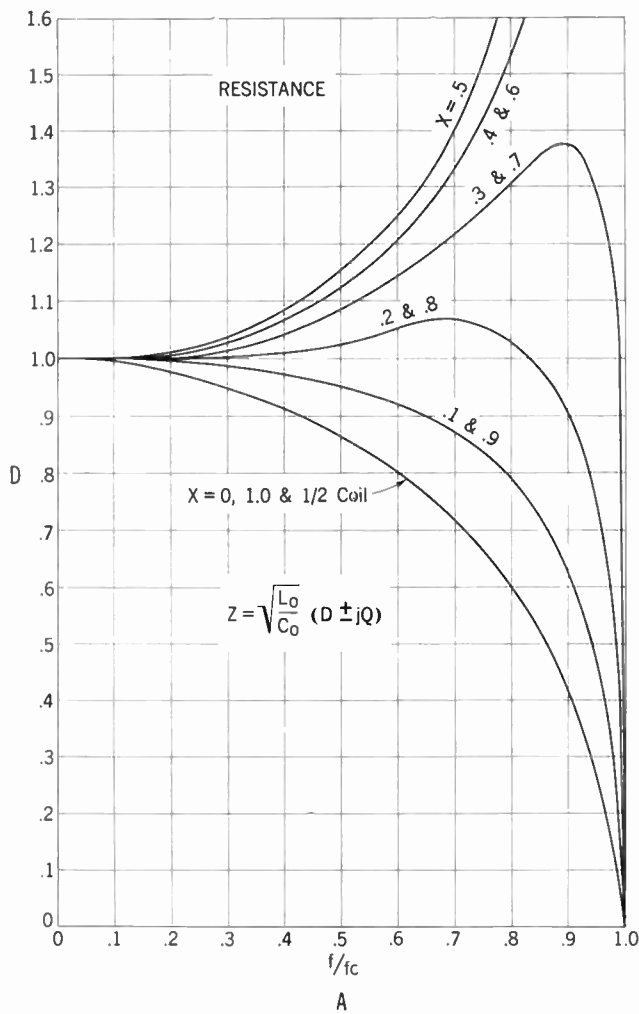
FIG. 28-3 IMPEDANCE COMPONENTS OF IDEAL LOADED CIRCUIT AT VARIOUS END SECTIONS

terminations, the frequency band being that up to and including the critical frequency. (The scale for frequency is shown as fractions of the critical frequency rather than as cycles in order that the curves may apply to any case.) Figure 28-3 (B) shows the corresponding reactance components.

An inspection of Figure 28-3 (A) shows that for a .2 or .8 section termination, a plain non-inductive resistance will approximate the resistance component of the circuit, as this resistance component remains nearly constant through the band of frequencies that the loaded circuit would be expected to transmit. This is true only for these two terminating conditions. Accordingly, if we choose the .2 section sending-end termination as that for which the basic balancing network is to be designed, we only need to connect in series with a resistance some combination of inductance and

capacitance that will approximate the corresponding reactance component shown in Figure 28-3 (B) in order to obtain a network which will simulate almost exactly the ideal loaded line; and, except at very low frequencies where the resistance of the actual line causes the impedance to depart appreciably from that of the ideal line, will closely approximate an actual loaded line. This combination is found to be a capacitance value in parallel with an inductance value. The essential elements of a simple basic network for a loaded circuit at .2 section termination are shown in Figure 28-4 (B).

The basic network is only intended to balance the characteristic impedance of a smooth line of infinite length in the case of open wire circuits, or an approximate .2 section termination sending-end impedance for a smooth line of infinite length in the case of loaded cable circuits. But the actual sending-end impedance of the circuit may vary

[270]
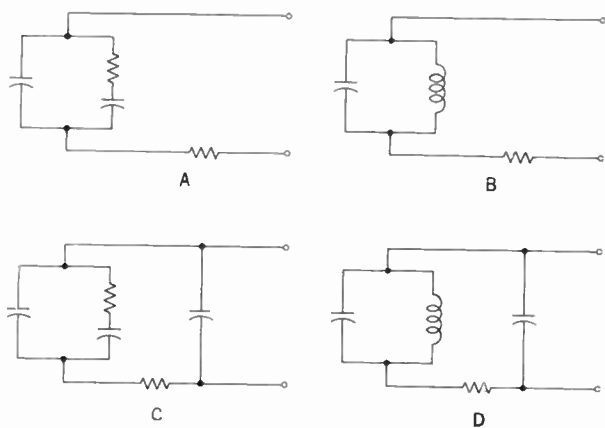
FIG. 28-4 BASIC BALANCING NETWORK CIRCUITS

widely from the particular impedance which the basic network is designed to balance. This may be due to the use of toll entrance cable in the case of open wire circuits, or to the fact that the termination may not be at the .2 section point in the case of loaded cable circuits.

It is the practice to make adjustments on the network sides of the telephone repeater's hybrid coils to take care of these irregularities by means of "building-out sections". That is, if an open wire circuit has a short section of non-loaded toll entrance cable, a capacitance value equal to the capacitance of this section may be bridged directly across the basic network as illustrated by Figure 28-4 (C). If the toll entrance cable is long, it may be necessary to compensate for its resistance as well as its capacitance and the building-out section may then include a series resistance as well as the bridged capacitance.

Similarly, in the case of the loaded cable circuit, if the capacitance on the office side of the last loading point is greater than that corresponding to .2 loading section, it is necessary to build out the basic network to adjust for this capacitance, as shown in Figure 28-4 (D). If, on the other hand, the circuit should be so terminated that the capacitance from the office side of the last loading coil was less than that of .2 loading section, it would be necessary to add bridged capacitance to the line of such value as to make the termination equivalent to .2 of a section.

In order to keep at a minimum the number of different basic network designs, it is current practice to employ networks whose characteristics can be varied somewhat by an appropriate strapping of certain terminals. Thus Figure 28-5 (A) shows the elements of a network designed for balancing

104 open wire side circuits. The additional terminals 3 and 4 may be strapped to terminal 2 so as to short out either resistance R_1 or R_2, or both. This makes it possible to adjust the network for different wire spacings of this particular type of facility. Similarly, the network of Figure 28-5 (B) represents the basic design for a 19-gage H-44-S cable circuit, and the strapping possibilities permit adjustments for deviations of the circuit capacitance from its nominal value. Both networks are made up with built-in building-out capacitors which can be adjusted to the desired values. Networks of like designs are available for other standard types of circuit facilities.

28.6 Transmission Regulation

The continued satisfactory operation of a long telephone circuit requires that the net overall loss remain approximately the same at all times. If there is no appreciable variation in the attenuation losses of the line sections between repeaters, the gains of the repeaters must be held constant; or, failing this, any variations in attenuation must be promptly compensated for by equal variations in repeater gains. Telephone repeaters and routine maintenance methods have been developed to a point where it is not difficult in practice to hold the gains of repeaters constant at any desired value. However, variations in the attenuation of line conductors due to temperature changes are inevitable.

The magnitude of net variation in total equivalent of a circuit, caused by temperature changes, is of course proportional to the total gross attenuation of the line circuit, since it depends on the variation in the resistance of the copper line wires. In open wire and aerial cable circuits, a daily change in resistance value of some 5 per cent, which corresponds to a temperature change
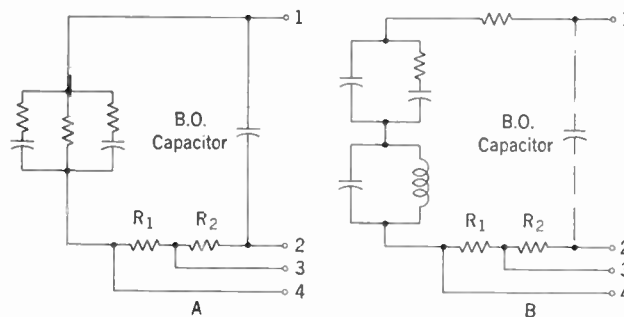
FIG. 28-5 ADJUSTABLE BALANCING NETWORKS

TABLE XIII

TRANSMISSION EQUIVALENTS IN DECIBELS PER MILE OF 16 AWG & 19 AWG CABLE CIRCUITS AT 55° F.
SHOWING YEARLY VARIATIONS IN EQUIVALENTS DUE TO TEMPERATURE CHANGES

| LOADING | TYPE | 16 AWG | | | | 19 AWG | | | |
| | | Side | | Phantom | | Side | | Phantom | |
		At 55° F.	Yearly Variation	At 55° F.	Yearly Variation	At 55° F.	Yearly Variation	At 55° F.	Yearly Variation
H-172-63	A.E.	16	± .018	16	± .018	28	± .031	28	± .032
	U.G.		± .006		± .006		± .010		± .011
H-88-50	A.E.	19	± .022	16	± .019	35	± .041	30	± .035
	U.G.		± .007		± .006		± .014		± .012
H-44-25	A.E.	25	± .029	21	± .024	17	± .055	39	± .046
	U.G.		± .010		± .008		± .018		± .015

Note: The loss at any other temperature, T, is approximately the 55° F. loss shown above plus the quantity $\dfrac{T-55}{55}$ times the Yearly Variation.

of about 22° F, may be expected. On a 1000-mile 165 open wire circuit, the total line attenuation of which is about 30 db, this would mean a variation in net overall equivalent of only about 1.5 db, assuming repeater gains to be held constant; on the other hand, 5 per cent of the total wire attenuation of a 1000-mile 19-gage H-44-S cable circuit, about 480 db, amounts to some 24 db, which is several times the value of the net equivalent of an average circuit. Variations over longer periods are of course much more severe. Table XIII shows the maximum yearly variations in equivalents per mile that may be expected in the more common types of cable circuits. It is obvious that it would be hardly possible to maintain service on long cable circuits without the aid of some automatic means of changing the gains of repeaters to compensate for changes in line attenuation due to temperature variation.

Long cable circuits are broken up into sections averaging about 150 miles in length, known as "circuit units". One of the repeaters in each circuit unit is a "regulating repeater", the gain of which is automatically changed by a "master regulator" in accordance with changes in temperature. The master regulator employs the principle of the balanced bridge. One arm of the bridge consists of a cable pair known as a "pilot wire", which extends through the same length of cable as the circuit units to be regulated. When the resistance of the pilot wire changes due to a temperature change along the cable line, the gal-

vanometer of the master regulator tends to deflect. By means of an auxiliary circuit, this causes a shaft driven by a small motor to turn until the bridge is again balanced. The rotation of this shaft also causes the operation of master relays which in turn control the operation of relays in all the regulated repeaters, causing the gains of these repeaters to be changed in proportion to the change in temperature of the pilot wire. This change in gain is accomplished by means of regulating networks or potentiometers, associated with the repeaters. In the case of 4-wire circuits, the regulating networks consist of potentiometers connected across the inputs of the 44-type repeaters (see Figure 25-2). In 2-wire repeaters, the regulating networks consist of artificial lines, or H-type pads, placed in the repeater circuit between the bridge points and the manual potentiometers (see Figure 25-1).

Regulating repeaters are ordinarily arranged to vary their gain in 1 db steps a total of plus or minus 10 db from their nominal designated values. For example, a 4-wire regulating repeater may be adjusted for a gain of 20 db at an average temperature of 55° F. and this gain may be automatically lowered to 10 db in cold weather or increased to as much as 30 db in hot weather.

The operation of the regulating system is illustrated by Figure 28-6. Here it will be noted that the bridge proper has two equal ratio arms, A and B, while the third arm, X, consists of the combination of the two sections of the pilot-wire

circuit in parallel (which are made equal at 55° F. by the adjustable resistances, S_1 and S_2) together with a part of the slide-wire resistor. In the fourth arm, C, is a fixed resistance, K, and the remainder of the slide-wire resistor. Since arms A and B are equal, balance of the bridge is secured when the total resistance of the third arm X and the fourth arm C are equal.

The regulating networks of the repeaters and the equalizing arrangements are designed on the basis of 55° F. being the average cable temperature. Consequently, the regulator is normally adjusted so that the slide-wire contact arm will be at mid-position (step 0) when the cable is at this temperature. Now assume that the temperature increases and thus increases the resistance of the pilot wire. Current then flows through the galvanometer and causes it to deflect. This will have no effect until the increase in temperature is great enough to cause an increase in the overall loop

resistance of the pilot wire of as much as 180 ohms. At this point, the deflection of the galvanometer becomes sufficiently large to cause movement of a mechanism to take place in such a way that the shaft on which the slide-wire arm is mounted is caused to turn in the proper direction to restore balance. When the arm has moved sufficiently to restore balance, current no longer flows through the galvanometer and the movement of the mechanism stops. Since the increase of 180 ohms in the total loop resistance of the pilot wire produces an increase of 45 ohms in the joint parallel resistance of the two sections, to restore balance it is evident that the slide-wire contact is required to transfer 22.5 ohms from arm X of the bridge in which the pilot wire is included, to arm C.

As the slide-wire contact transfers 22.5 ohms from one arm of the bridge to the other, the brush arm, F, moves from one stud on the dial switch to
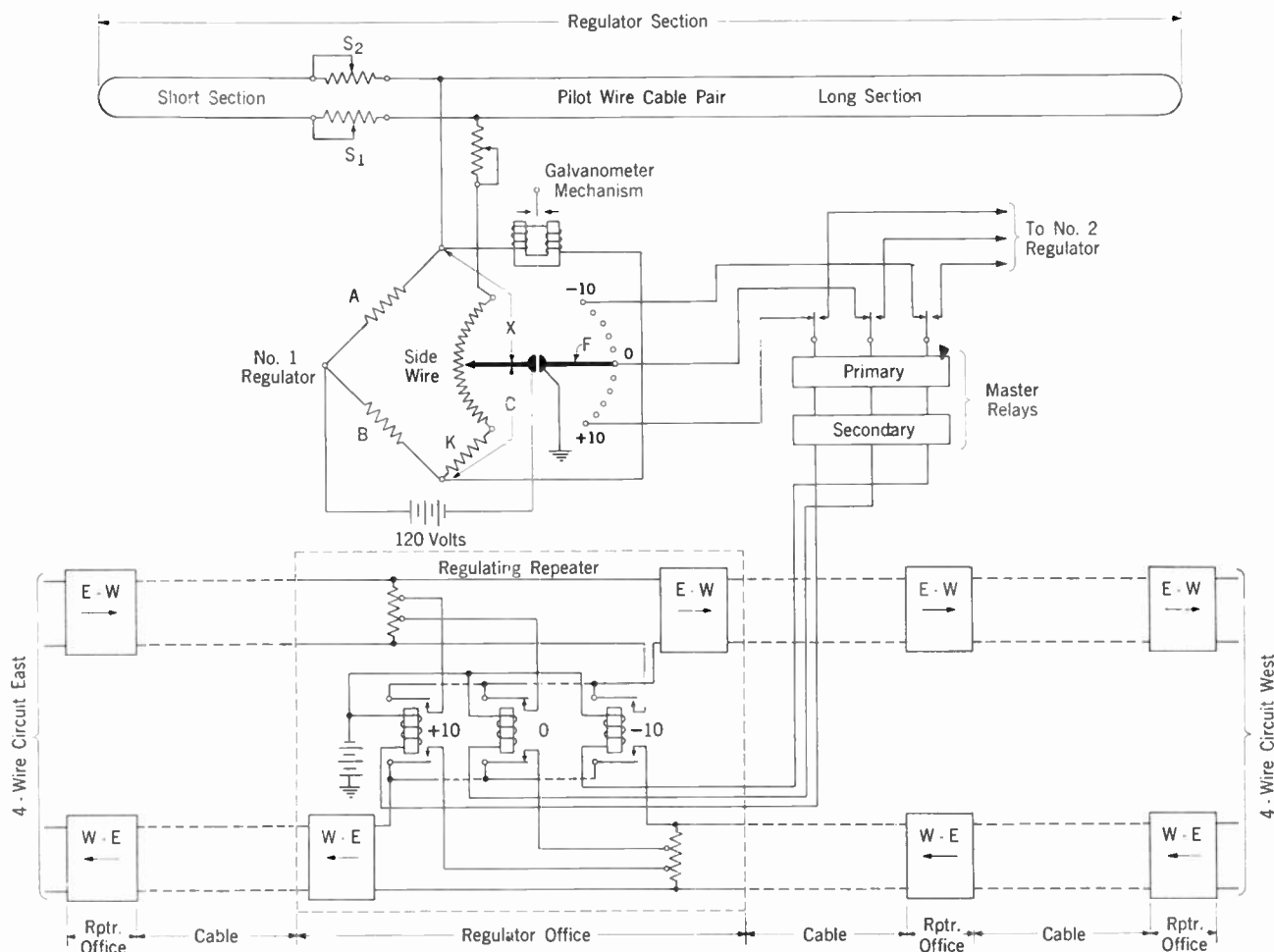
FIG. 28-6 PILOT WIRE TRANSMISSION REGULATOR CIRCUIT

[273]

the next adjacent stud. This movement is spoken of as a movement of one step and, as is evident from the above description, it corresponds to a change of 180 ohms in the overall loop resistance of the pilot-wire circuit. There are 21 studs, consisting of a zero step and 10 steps each side of the zero step (only 10 steps shown in Figure 28-6). The movement of the brush arm causes the master relays to operate and so adjust the regulating networks of all the regulating repeaters.

In Figure 28-6, the brush arm of the slide-wire mechanism is shown on the 0 stud of the dial switch, and ground is then connected to the master relays. As a result, these relays are so operated that the center relay in each regulating repeater is likewise operated. The latter relays are then connecting the proper resistances in the input potentiometers of the repeaters to provide the prescribed gain for the 55° F. temperature. When the brush arm moves a step, other master relays are operated and, in turn, other relays in the regulating repeaters. These change the potentiometer resistances in such a way as to effect the appropriate change in gain.

One master regulator is capable of controlling a large number of regulating repeaters. However, since the change in the loss of a circuit with varying temperature depends upon the gage of conductors and type of loading, the proper gain variations for a given temperature change may not be the same for all the circuits under the control of a single master regulator. In order to use the same master regulator under these conditions, the systems are designed so that a given movement of the master regulator produces different changes in the gains of the regulating repeaters on the different types of facilities. This is the purpose of the two sets of master relays, primary and secondary. A separate chain of secondary relays is provided for each of the different types of regulating networks. The primary relays are operated directly from the dial switch on the master regulator. These relays are numbered from +10 to zero to −10, corresponding to the stud from which they are operated. The contacts on the primary relays are used only to operate the secondary relays, and do not control directly any of the regulating repeaters. Each of the different groups of secondary relays then controls directly the relays of all of the like regulating repeaters with which it is associated.

Perfect compensation for temperature changes requires that the pilot wire be loaded with the same kind of coils as the transmission circuits which are to be regulated. This would, in general, require separate regulators and pilot wires for sides and phantoms, as well as for the different types of loading and for the different gages. Since 19-gage H-44-25 4-wire circuits require the most accurate regulation (due both to the greater lengths for which they are used and to the fact that the variation per mile is greater than for most other loading now in use), the pilot-wire regulating system is designed on the basis that 19-gage H-44-25 loaded pairs will be used for pilot wires. The steps on regulating repeaters for other types and gages of circuits are laid out in such a way that substantially accurate regulation is obtained with these pilot wires.

Figure 28-7 is a transmission level diagram of a 4-wire circuit made up of three regulator sections or circuit units. It will be noted that at all repeater stations except those at the terminals of the circuit units, the input and output levels vary with temperature. It follows that it is necessary to know the setting of the regulating repeater in order to know the proper levels at any of these repeaters at any particular time. Levels at the ends of the units are constant, however, under the normal condition where each circuit unit is completely regulated by the regulating repeater that it includes. Regulating repeaters are usually located at the repeater station nearest the midpoint of the circuit units, although this is not strictly necessary in every case.
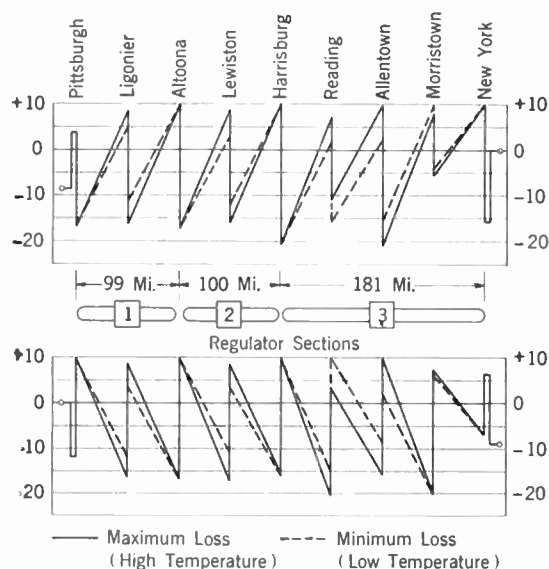
FIG. 28-7 ENERGY LEVEL DIAGRAM FOR FOUR-WIRE CABLE CIRCUIT

[274]

A

B
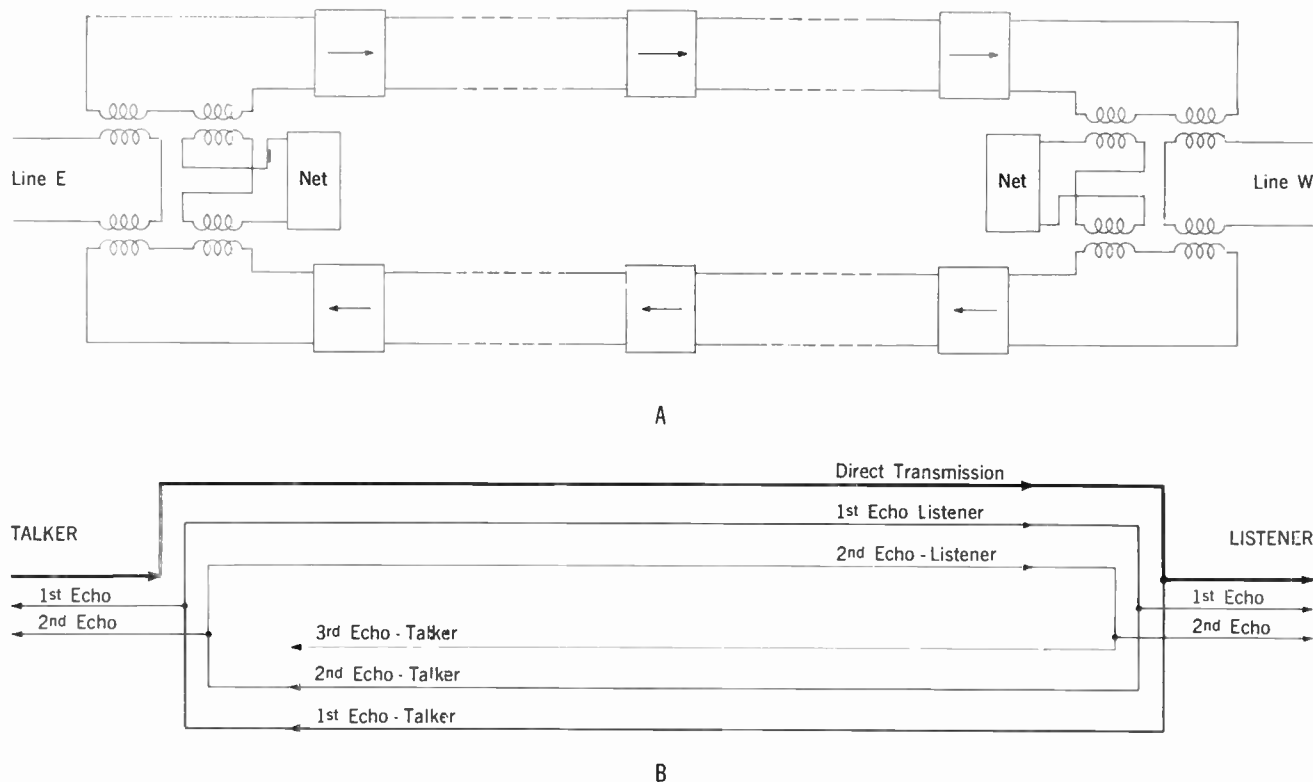
FIG. 28-8 ECHO PATHS IN FOUR-WIRE CIRCUIT

28.7 Echo Control

Another series of problems, largely peculiar to the longer cable circuits, arises from the fact that the velocity of propagation over such circuits, as now loaded, is relatively low. The time required for transmission over circuits of this type may thus become quite appreciable. If when a conversation is being carried on, some portion of the speaker's voice is returned toward him from a sufficiently distant point, the effect will be like that of an ordinary echo. This will obviously be rather disconcerting to the speaker if the time factor is great enough so that he hears entire syllables repeated back to him. In any case, such an echo effect tends to degrade the quality of transmission and it must be guarded against in long, low-velocity circuits.

Figure 28-8 (A) shows schematically a long 4-wire circuit layout. When the person at the east terminal is talking, the voice currents are sent through the 4-wire terminating set to both sides of the circuit. Transmission over the lower (receiving) side stops at the output side of the terminal repeater, but the voice currents are transmitted over the upper side until the 4-wire terminating

set at the west terminal is reached, where the energy is divided between the network and the 2-wire line connected at that terminal. This transmission is indicated by the heavy line in Figure 28-8 (B) marked "direct transmission". With perfect balance between the network and the 2-wire circuit at the west terminal, no further effects will be produced.

As we have already noted, however, there is always some unbalance at the 4-wire terminating sets and, consequently, a small current passes into the lower branch of the circuit at the west terminal and is propagated back to the talking station at the east terminal, as is indicated by the line marked "1st echo—talker". This is heard at the east terminal either as side-tone, or as a distinct echo if the time of transmission around the circuit is great enough. Such currents are called unbalance or echo currents affecting the talker.

Due to unbalance at the east terminal, another current, derived from the first echo affecting the talker, is propagated from the east to the west, forming another echo which is heard by the listener. This is indicated by the line marked "1st echo—listener" in Figure 28-8 (B). Such currents are called unbalance or echo currents affecting the

[275]

listener. The first echo current affecting the listener through the unbalance at the west terminal gives a "2nd echo—talker" current at the east terminal, and this action may go on indefinitely.

If the total loss around the unbalance current path, including the loss through the 4-wire terminating sets, is greater than the total gain of the repeaters, the successive echoes die out rapidly. With a small margin, there may be several echoes of sufficient magnitude to affect the persons at each end of the circuit, and if the loss becomes equal to or less than the gain, the circuit will sing.

If the transmission to and fro in a circuit could be accomplished instantaneously and the unbalance did not vary with frequency, the waves constituting the echoes would occur simultaneously with the waves of the direct transmission and would tend either to re-enforce or weaken the direct transmission, depending upon the poling around the unbalance path. The net results of the echo currents would then be merely to increase or decrease the transmission equivalent between the east and west terminals and to introduce more or less side-tone at both terminals.

However, as shown in Table X, the velocity of propagation of the various types of facilities most commonly employed for voice-frequency cable circuits is actually less than 20,000 miles per second, and an appreciable time is therefore required for propagation over the longer circuits. Each successive echo accordingly arrives after a definite time interval, depending upon the length of the circuit and the velocity of propagation. The interfering effect depends on both the volume of the echo and the time-delay. A given volume of echo produces a greater disturbing effect as the time-delay increases and vice versa.

While the 4-wire circuit offers only one round-trip path for echo currents, a 2-wire circuit involves a large number of such paths. If the sections of line between repeaters were perfectly constructed, if the impedance of the repeaters were such as to form perfect terminations for the sections of line, and if the networks balanced these perfect lines perfectly, transmission could take place from one end of the circuit to the other without setting up any unbalance currents except those reflected from the terminals. Such an ideal circuit would give the same performance as a 4-wire circuit with equal terminal unbalances.

In practice, however, there is more or less unbalance on each side of each repeater so that, as the direct transmission passes through each successive repeater, an unbalance current is set up which travels back toward the talker, giving an echo current for each repeater in the circuit. Each of these first echo currents in turn sets up an echo current traveling toward the listener at every repeater that it encounters, and each of these echo currents sets up another echo current at each repeater it encounters. This process continues indefinitely but the successive echo currents are attenuated rapidly to inappreciable magnitudes. The action, however, does give rise to a very large number of echo currents. The final effect upon the talker or listener naturally depends upon the resultant of all these currents.
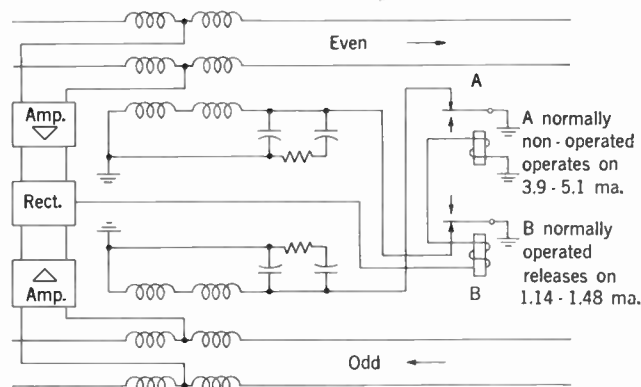
FIG. 28-9 PRINCIPLE OF ECHO SUPPRESSOR

These unbalance currents increase in volume as the overall net equivalent of a circuit is decreased, due to raising the repeater gains. For each circuit, therefore, there will be a certain minimum permissible net overall equivalent because of the unbalances which are present in the circuit. Any further increase in the repeater gains will cause the unbalance currents to become too large to be tolerated. Furthermore, since each additional repeater in the circuit increases the number of unbalance paths, the total echo effect tends to limit the total number of repeaters that can be operated in tandem in a 2-wire circuit having a practicable overall net equivalent.

On the other hand, in the case of 4-wire circuits the possible minimum net equivalent, in so far as echo effects are concerned, is limited only by the extent of the unbalance at the two terminals. If the minimum net equivalent is still too high, echo suppressors may be inserted in the circuit to break the echo paths and thus permit reducing the overall equivalent to the desired value.

The principle of operation of the echo-suppressor may be understood by referring to the simplified schematic of Figure 28-9. As there indicated, hybrid coils are inserted in each side of a 4-wire circuit or in each side of one of the repeaters of a 2-wire circuit. The bridge points of the hybrid coils are connected to separate amplifiers, the outputs of which connect to a common rectifier circuit. The rectifier output is fed in series through the windings of relays B and A, as shown. The rectifier circuit is designed to act differentially. When there is no energy, or only low levels of speech or noise in both sides of the 4-wire circuit, the d-c rectifier output current will be at its normal value which lies between 3.9 ma, the minimum operating current for relay A and 1.48 ma, the maximum release current of relay B. Relatively high speech energy in the even side of the circuit will cause the rectified current to increase above normal value; and high speech energy in the odd side causes the rectified current to decrease below normal. In the first case, relay A operates while B is unaffected and in the second case, relay B releases while A is unaffected. Operation of relay A removes the ground connection which had been short-circuiting the resistance capacitance network connected to the series winding of the odd hybrid coil. The impedance of this network balances, in the voice-frequency range, the input impedance of the amplifier bridged across the hybrid coil. When the short-circuit is removed from the network, accordingly, the now balanced hybrid coil inserts a large loss in the odd path, thus effectively suppressing energy transmission over that side of the line. Similarly, when the B relay releases as a result of high energy in the odd path, the ground connection is removed from the even network and transmission over that side of the line is suppressed.

29.1 Types of Carrier Systems

The general principles of carrier operation were outlined in Chapter 27 and these principles apply alike to all of the several types of carrier systems used in the telephone plant. With respect to the frequency allocations employed, the systems in use range all the way from the voice-frequency carrier telegraph system, which operates entirely within voice range, up to systems applied to co-axial conductors where frequency bands millions of cycles in width may be involved. As has been previously noted, radio transmission systems have much in common with carrier systems in that both make use of the frequency-division principle for obtaining numbers of separate transmission paths over a single transmission medium. Radio systems as employed in telephone work, however, are discussed in other Chapters. In general, the term carrier system is here used only with reference to frequency-division multiplex systems employing as a transmission medium physical facilities such as open wires, cable pairs, or coaxial conductors.

Figure 29-1 shows the overall frequency allocations for the types of telephone and telegraph carrier systems currently in use in the Bell System. In most of the carrier systems applied to open wire conductors (Type-B telegraph and Types C, J and O telephone) several different sets of frequency allocations are used within the over-all range indicated in Figure 29-1. The purpose of this is to reduce the possibilities for inter-system crosstalk where a number of systems are operated on the same pole line. All but one (Type-G) of the open wire carrier systems use separate channels for transmission in opposite directions, thus in effect operating on a 4-wire basis. The Types K and N systems, which are designed for application to wire cable facilities, employ the same channel frequencies for transmission in both directions but use two pairs of conductors, one transmitting in each direction. This is also true of the voice-frequency telegraph system, which either operates on a 4-wire voice-frequency cable circuit or on the two one-way channels of an open wire or cable carrier tele-

phone circuit. Type-L systems, which are applied to coaxial conductors, also operate on a 4-wire basis.
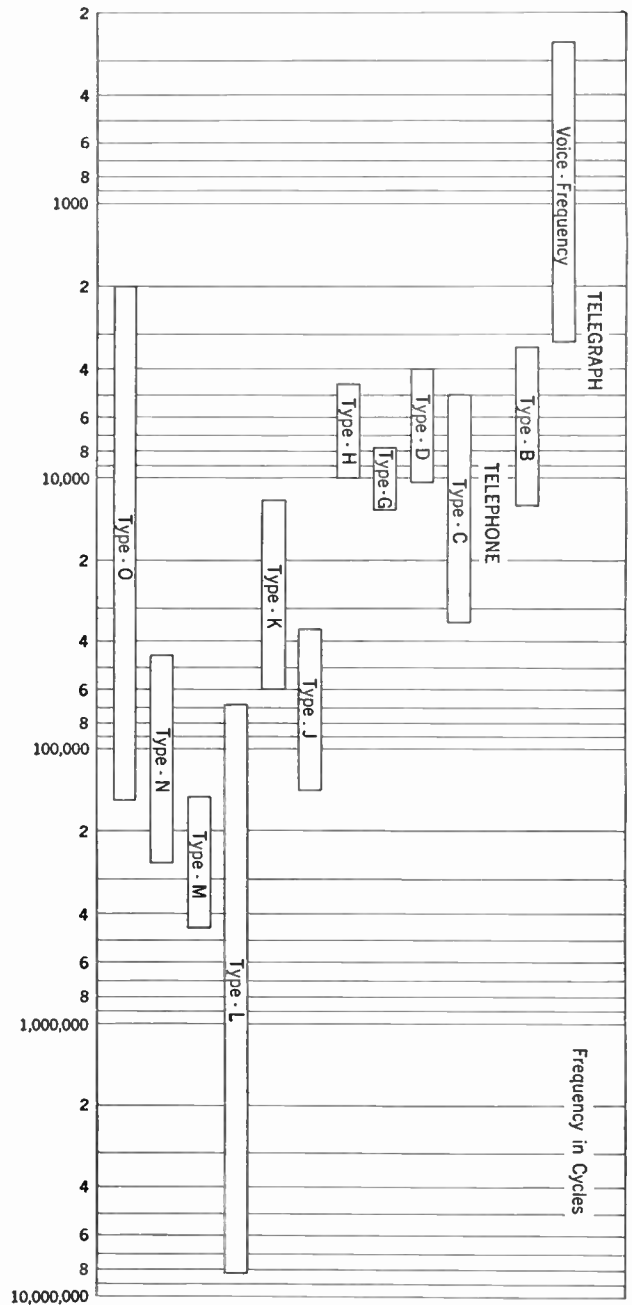
FIG. 29-1 OVERALL FREQUENCY ALLOCATIONS OF STANDARD TYPES OF CARRIER SYSTEMS

The principles of most of the major apparatus units and circuit arrangements that are employed to make up a carrier system such as filters, equalizers, modulators, amplifiers, etc., have been discussed in earlier Chapters. This and the following Chapter will be devoted to a brief survey of the general overall design of the types of systems most widely used in long distance telephone and telegraph work. It will include discussion of the carrier frequencies used to obtain the various channels, the modulation steps required, and the important regulating and equalizing procedures that must be employed for satisfactory overall transmission.

29.2 Carrier Telegraph Systems

As noted in Article 27.2, a simple method of carrier telegraph operation depends on transmitting spurts of alternating current of a particular frequency to represent marking signals and cutting off this current for spacing signals. Simultaneous operation of a number of telegraph channels over a single facility is then effected by using a different a-c frequency for each channel. Since each channel is distinguished by its own single frequency, only relatively simple selecting circuits need be used at the terminals to secure effective channel separation. This is illustrated by Figure 29-2 which diagrams one of the earliest carrier telegraph systems that permitted up to ten two-way telegraph channels to be superimposed on an open wire telephone circuit. As indicated, it employed twenty carrier frequencies in the range between 3330 and 10,000 cycles per second. The ten lower frequencies were used for transmission in one direction and the ten higher frequencies for transmission in the opposite direction. This permitted ready separation of the two directions of transmission at terminals and repeater points by means of a pair of simple directional filters, one of which passed only frequencies above 6000 cycles and the other only frequencies below that value.
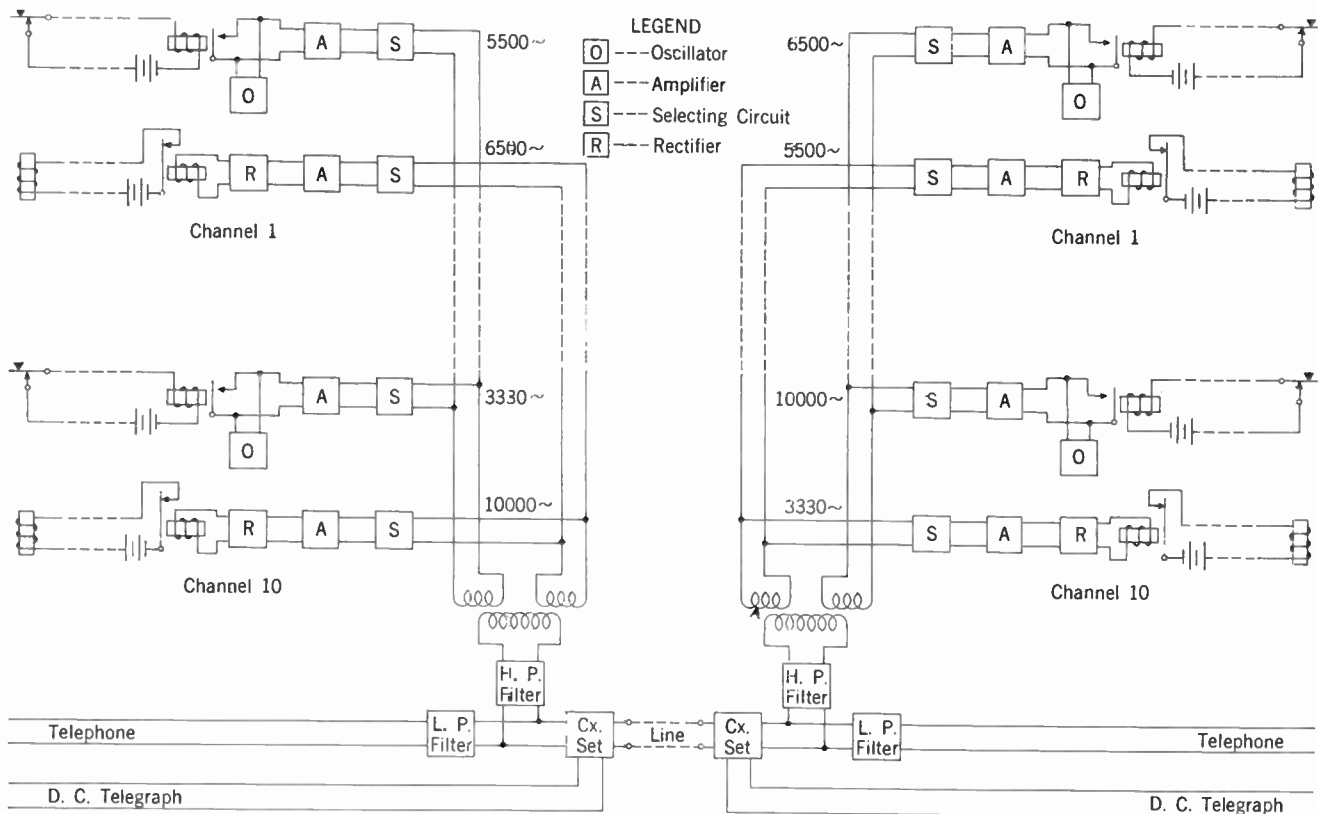
FIG. 29-2 HIGH-FREQUENCY CARRIER TELEGRAPH SYSTEM

Carrier telegraph systems of the type discussed above have very limited application in current long distance telegraph practice because the frequency space which they would occupy can usually be employed to better advantage for carrier telephone systems. They have been generally superseded by the so-called voice-frequency systems which operate on a 4-wire cable circuit or on a voice channel of a carrier telephone system. The facility so used, however, cannot be simultaneously employed for telephone service because the telegraph carriers are within the voice band.

The original design of voice-frequency carrier telegraph system (40-Type) which is still extensively used, operates on essentially the same principles as the high-frequency system previously discussed. Marking and spacing telegraph signals are represented by the presence or absence of current. Since the transmission medium is either actually or effectively a 4-wire circuit, there is no problem in separating the transmitting and receiving channels. The same carrier frequencies are used for transmission in both directions. These systems provide up to 18 two-way telegraph circuits. The carrier frequencies employed are odd multiples of 85 cycles beginning with channel 1 at 425 cycles and extending at 170-cycle intervals through channel 17 at 3145 cycles. Channel 18, when used, is placed below channel 1 at 255 cycles. Because of the generally poor transmission characteristics of line facilities at that frequency, however, this channel is likely to be inferior to the higher frequency channels.

The carrier sending and receiving apparatus is similar in principle to that of the older high-frequency system. Spurts of the carrier current are sent over the line as the transmitting amplifier (modulator) is shorted out by the telegraph impulses, and a rectifying device (demodulator) at the receiving end converts the spurts of carrier back again to ordinary d-c telegraph signals. The channel receiving circuit includes a two-stage amplifier, a bridge-type varistor rectifier, an output and control tube, and a receiving telegraph relay, which is operated by the plate current of the control tube and is normally biased to spacing by an auxiliary winding. The first stage of the amplifier is arranged to provide partial automatic gain control. The gain of the tube in this stage is variable, and the circuit is designed so that the amount of gain varies in approximate inverse proportion to the amplitude of the incoming signal voltages. The voltages impressed on the rectifier, and the resultant rectified signal, are therefore nearly constant in magnitude regardless of the input level. The control tube acts effectively to amplify the rectified signals and to "square-up" the signal form. In other words, the tube circuit is so arranged that during spacing intervals, when there is no incoming rectified signal voltage, its control grid is biased well beyond the cutoff point and no current flows in the plate circuit or the winding of the receiving relay to which the plate is connected. During marking intervals, on the other hand, the rectified signal voltage overcomes the bias so that a large plate current flows and causes the relay to operate sharply to its marking contact.
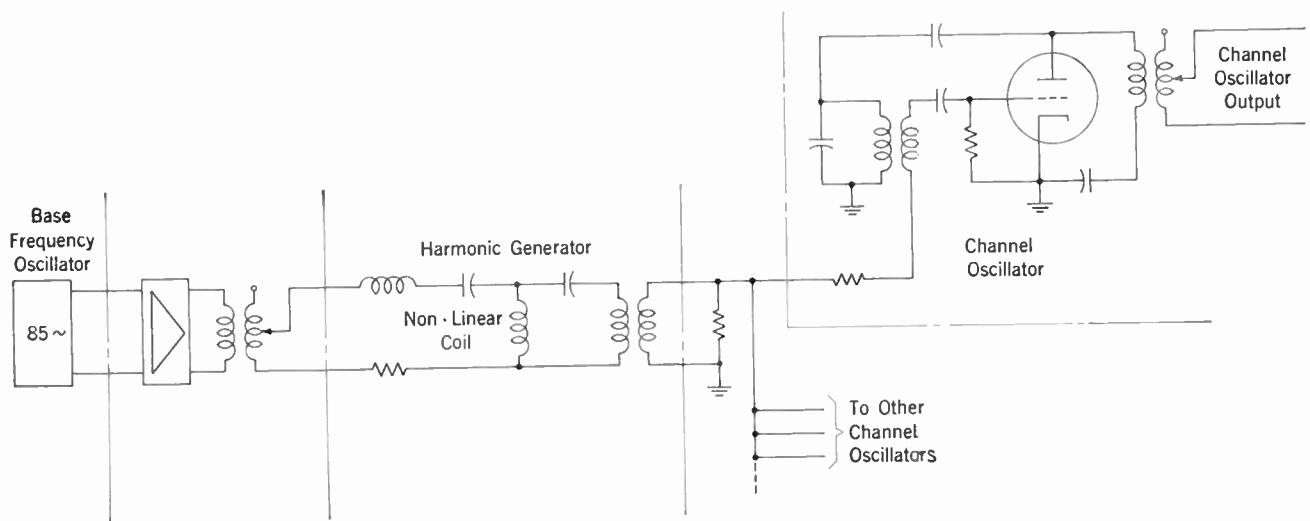
FIG. 29-3 V-F CARRIER GENERATING CIRCUIT
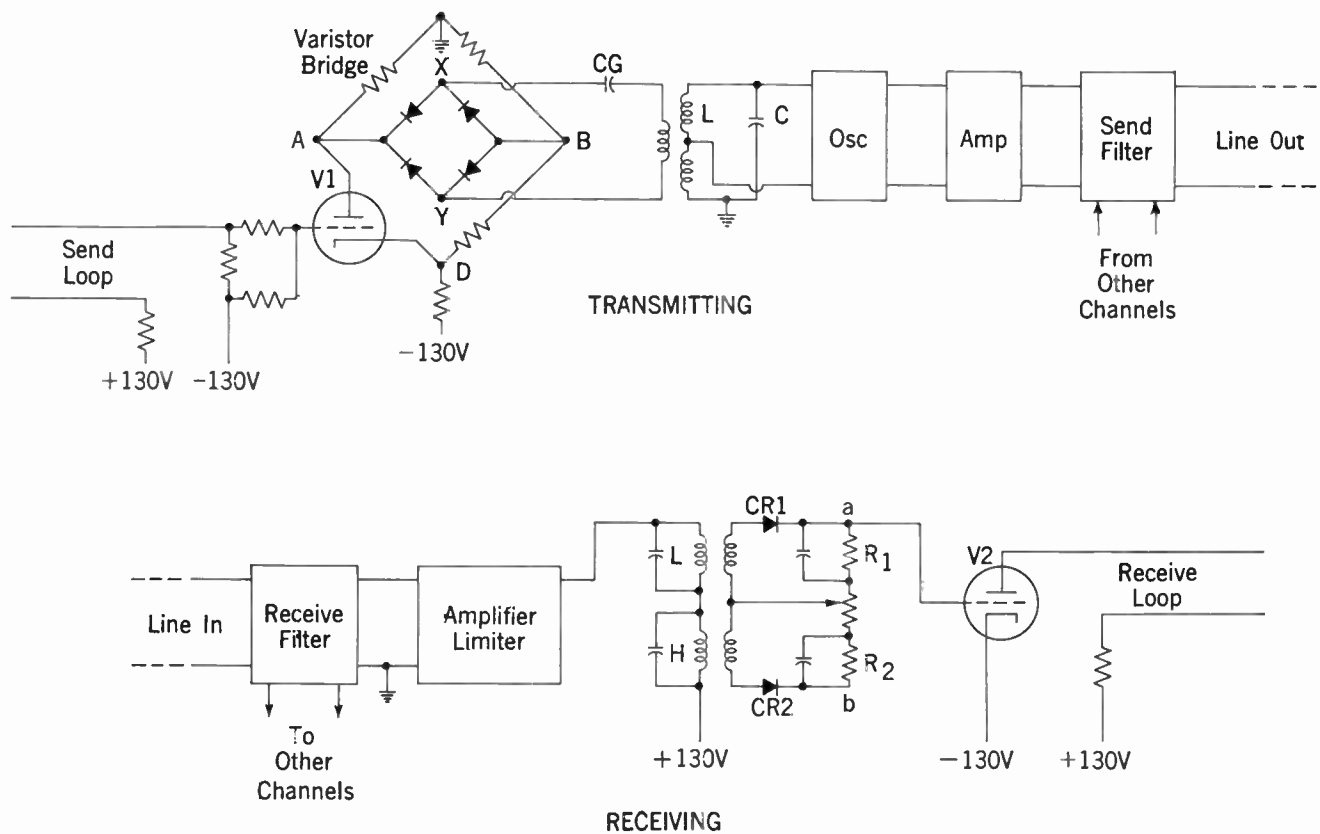
FIG. 29-4 FREQUENCY-SHIFT CARRIER TELEGRAPH SYSTEM

The carrier supply voltages in recent designs of voice-frequency telegraph systems are provided by vacuum tube oscillators in the manner indicated schematically in Figure 29-3. As shown, a base frequency of 85 cycles is developed by a common oscillator, amplified and passed on to a harmonic generator circuit. This includes a non-linear coil, or saturable reactor, connected in an arrangement like that described in Article 26.5. As was there pointed out, the output of such a circuit arrangement is a sharply peaked wave form which contains all the odd harmonics of the applied sine wave. This wave is introduced into the grid circuits of all the channel-frequency oscillators. Each channel oscillator tuned circuit then acts as a filter to select the particular harmonic to which it is tuned, and the channel oscillator falls into exact step with this frequency.

More effective in many respects is a more recent design of voice-frequency carrier telegraph system, coded 43A1, which employs a frequency shift method for distinguishing between marking and spacing signals. In this system, the marks and spaces are made by shifting the frequency 35 cycles up and down from the nominal channel carrier frequency. In contrast with the older systems where mark signals are produced by transmission of the carrier current and space signals by no current, the frequency shift method permits a continuous flow of current that does not change in value with the signal transitions. This makes for stable operation since the transmission conditions are the same at all times.

The method of operation of the system may be followed with the aid of the simplified sketches of Figure 29-4, which show the essentials of the terminal transmitting and receiving circuits. The circuits in this illustration are connected for low-frequency marking and high-frequency spacing. At the transmitting end, closing the sending loop for a marking signal impresses a positive voltage on the grid of vacuum tube V1. This makes the tube conducting so that the arm AD of the varistor bridge is effectively short-circuited. The potential at A is then lower than that at B and current flows through the four varistors in their forward di-

[281]

rection. In this condition the impedance of the varistors is very low and points X and Y are effectively shorted together. A circuit through capacitor CG is thus closed, which effectively inserts its capacitance into the LC circuit of the oscillator. The values of L and C are such as to give the oscillator an output frequency somewhat higher (35 cycles) than the nominal carrier frequency of this particular channel. Addition of the CG capacitance shifts the output frequency to a like value below the carrier frequency. This output is then amplified in a single stage and passed to the sending filter where it is joined by the outputs of the other channels for transmission over the line.

To transmit a spacing signal, the sending loop is opened and the grid of V1 is driven negative as the positive voltage is removed. The tube is then non-conducting and the effective open circuit causes the potential at A to be positive with respect to B. The voltage applied to the varistors is now in the reverse direction which causes them to present very high impedances. The impedance across XY is accordingly so high as to constitute an effective open circuit. This removes the capacitor CG from the tuning circuit of the oscillator which causes its output frequency to shift to the value above the carrier frequency, as determined by L and C alone.

The arrangement for discriminating between the marking and spacing frequencies at the re-

ceiving end of the line is indicated in the lower part of Figure 29-4. Here the input from the line goes first to a receiving filter where the several channels are separated by appropriate tuning elements. Each individual channel is then fed to a three-stage amplifier which is also designed to cut off any incidental current peaks that could affect the signal reception adversely. The a-c output of the amplifier flows in series through the two LC circuits designated L and H. The L circuit is sharply tuned for resonance at the low marking frequency while the H circuit is similarly tuned to the high spacing frequency. A received marking signal, accordingly, will establish a relatively high voltage across the resonant L circuit while having little effect on the H circuit. The secondary current resulting from the high voltage across L is rectified by the varistor CR1 so as to flow in only one direction through the resistance R_1. This causes a substantial voltage drop across R_1, which is positive at a and negative at b. The positive voltage applied to the grid of V2 makes the tube conducting, which closes the receiving loop circuit to battery for a marking signal.

Similarly, a high frequency spacing signal from the line develops a high voltage across the resonant H circuit. This in turn produces a large IR drop across R_2 which is positive at b and negative at a. The grid of V2 is accordingly negative, the tube is non-conducting, and the receiving loop is opened for a spacing signal.
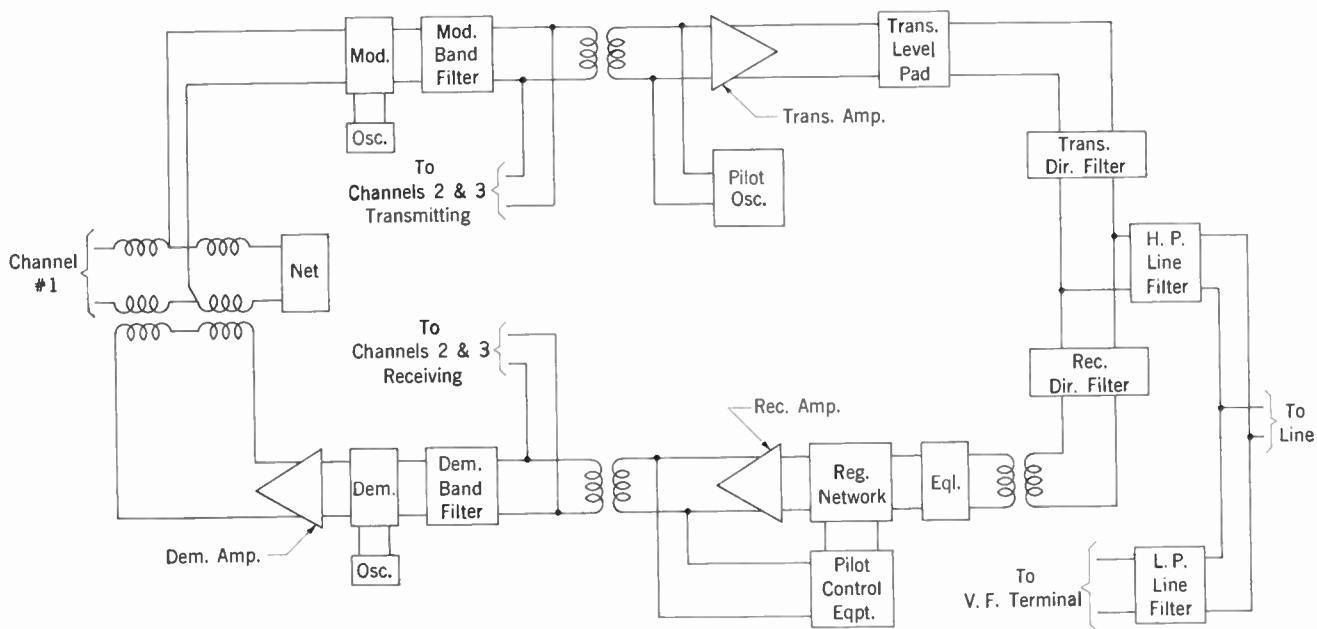
FIG. 29-5 TYPE-C CARRIER TELEPHONE TERMINAL

29.3 Type-C Carrier Telephone System

The oldest of the carrier telephone systems still in general use is coded Type-C. It operates on open wire facilities and provides three telephone circuits additional to the normal voice-frequency circuit. The general layout of the Type-C carrier terminal is shown schematically in Figure 29-5. Transmission over each channel consists of a single side-band, the carrier frequency being suppressed in the modulator. Earlier designs of C-systems used vacuum tube modulators and demodulators. Copper-oxide varistors are employed for both purposes in the more recent designs. The required carrier frequency at each end of each channel is supplied by individual vacuum tube oscillators whose stability is such as to maintain satisfactory frequency synchronization at all times.

The individual channel carrier frequencies used in Type-C systems are shown in Figure 29-6. It will be noted that a total of six different frequency allocations are used. All six allocations, however, cannot be used for any one C-system design. Allocations coded CN, CT, CS and CU are suitable
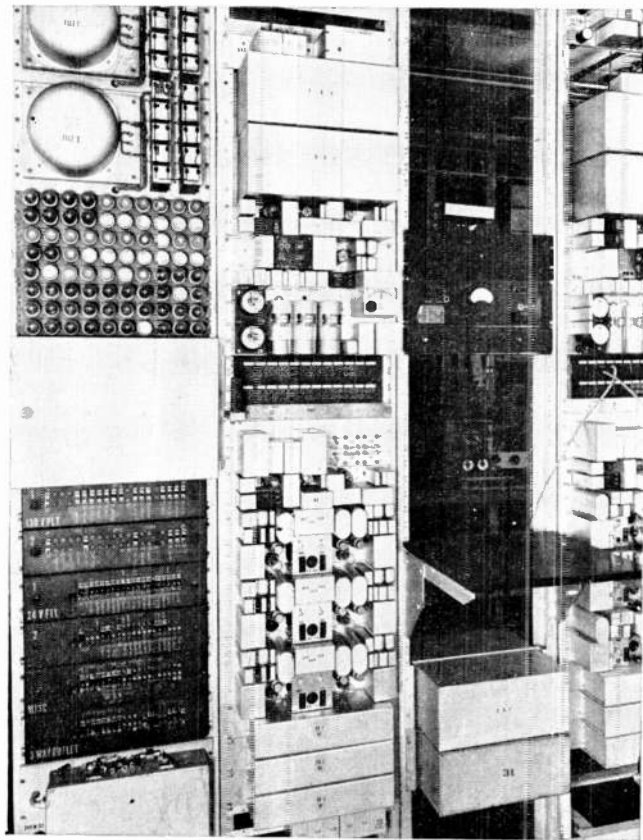
TYPE-C CARRIER TERMINAL

for the earlier designs of C-systems, while allocations CS, CU, CA and CB can be used with the more recent design. In all cases, the separation between carriers is at least 3000 cycles, which permits transmission of a side-band about 2500 cycles in width. Separation between the transmission in the two directions at terminal and repeater points

Note Arrows indicate carrier frequencies

| Upper side band | Lower side band
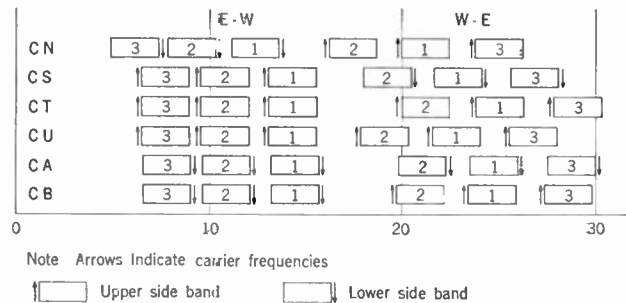
E-W

	3	2	1	PILOT
CN	7.6	10.6	13.9	10.55
CS	6.3	9.4	12.9	9.45
CT	6.3	9.4	12.9	9.45
CU	6.3	9.4	12.9	9.45
CA	9.3	12.4	15.9	12.35
CB	9.3	12.4	15.9	12.35

W-E

	2	1	3	PILOT
CN	16.1	19.8	23.7	19.85
CS	20.7	24.4	28.4	24.35
CT	19.8	23.7	27.7	23.75
CU	17.7	21.4	25.4	21.45
CA	22.5	26.20	30.2	26.15
CB	19.5	23.2	27.2	23.25

FIG. 29-6 TYPE-C CARRIER FREQUENCY ALLOCATIONS

is obtained by directional filters. This permits the use of a single amplifier common to all three channels transmitting in the same direction, at both terminal and repeater points. High-frequency signaling is employed, the ringing current being transmitted over the system in exactly the same way as the voice currents. C-systems require the use of repeaters at intervals of about 140 to 180 miles, depending upon the transmission characteristics of the conductors. The repeater consists of two amplifiers, one transmitting in each direction, together with directional filters for obtaining the necessary separation.

Regulation to compensate for changes in line attenuation due to weather and temperature changes is effected automatically by means of automatic

gain control circuits associated with the line amplifiers at repeaters and with the receiving amplifiers at terminals. The regulating equipment is controlled by pilot channel frequencies, which are supplied by an oscillator at the transmitting terminal as indicated in Figure 29-5. The values of the pilot channel frequencies for each direction of each system frequency allocation are listed in Figure 29-6. The pilot frequencies in each case have values lying between the carrier frequency and the side-band of the middle channel. Since the pilot frequencies are transmitted along with the regular signal frequencies, they suffer essentially the same attenuation losses in passing over the line and are affected by any change in the line characteristics in the same way, and to the same extent, as the message currents in their several channels. By establishing normal values for the pilot frequency currents at each amplifier point, therefore, any change due to changing line conditions may be caused to register in such a way as automatically to produce a correcting adjustment.

As shown in Figure 29-5, the pilot channel voltage is applied at the input of the terminal transmitting amplifier. At repeater points, this single-frequency voltage is tapped off at the output of the line amplifier and fed back through a rectifier circuit to a pilot channel control circuit which controls the net gain of a regulating network and amplifier connected into the main transmission path in front of the line amplifier. The same general plan applies at the receiving terminal, except that here the pilot channel voltage is taken off at the output of the receiving amplifier.

The regulating amplifier circuit consists essentially of a variable attenuator in series with a vacuum tube amplifier. The attenuator, known as the regulating network, is designed to have loss-frequency characteristics similar to those of the line and is divided into three units of equal loss. Its net loss to through transmission is varied by means of a capacitor whose movable plate is rotated under the control of the pilot channel voltage. The rotor of the capacitor is connected to the
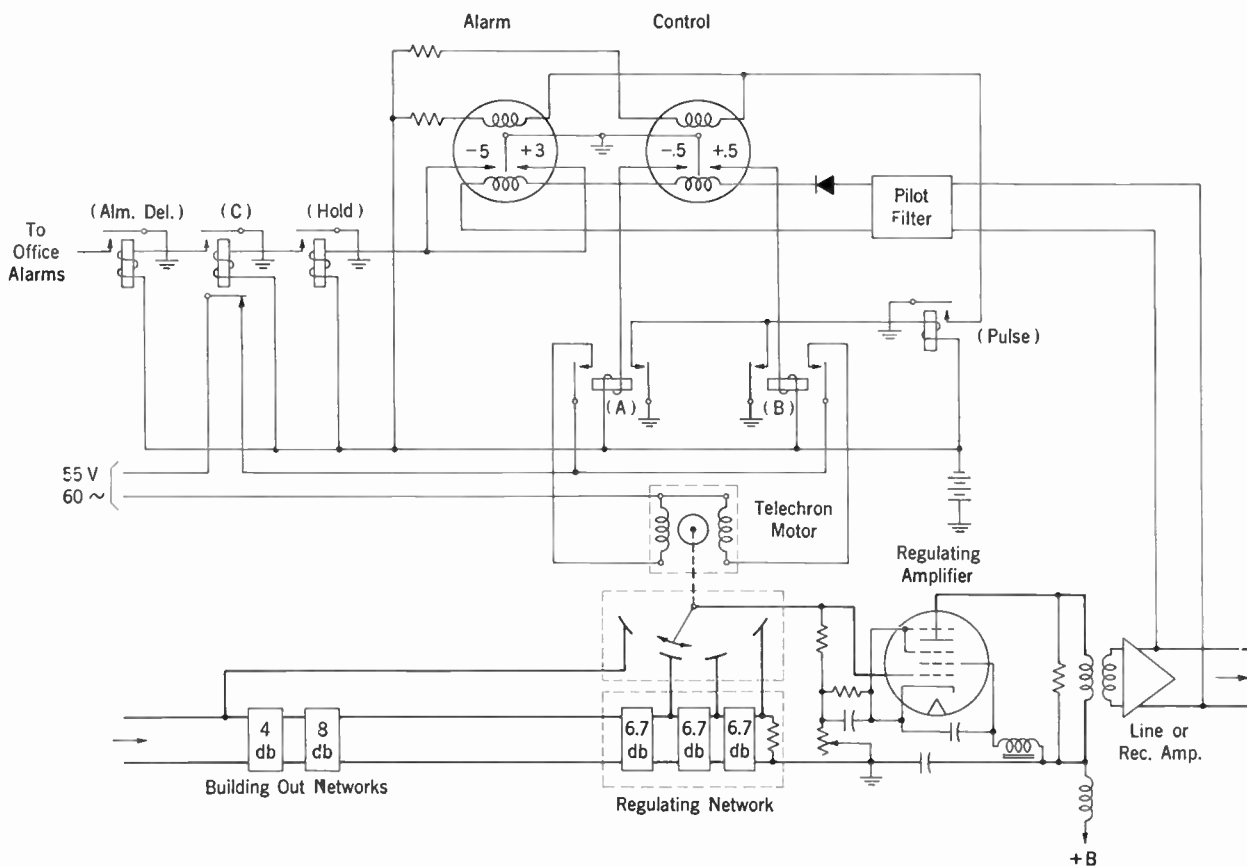
Fig. 29-7 C-Carrier Automatic Gain Control Circuit

control grid of the amplifier tube so that the voltage applied to the amplifier depends upon the position of the capacitor plate. When the capacitor rotor is at its extreme left position, the regulating network is effectively out of the transmission path. At its extreme right position, on the other hand, the entire network is in the transmission path; and at any other position, the loss inserted is some definite fraction of the total loss of the network. This arrangement provides a smooth control of the net gain of the regulating circuit and avoids the use of sliding contacts or relays in the transmission path.

Figure 29-7 shows how the pilot channel voltage controls the position of the regulating capacitor. As may be seen, the incoming pilot channel voltage is selected by a band filter, rectified, and led through the winding of a control relay. This latter is a highly sensitive type of relay designed to act very positively by means of the attraction of magnetic material on its armature to a small magnet on each of the fixed contacts. The relay is given a mechanical bias so that its armature is centered between the contacts when the rectified pilot current is at the normal level. Deviations from this normal level cause the armature to move sharply to one or the other of the contacts, where it will stay until released by the action of the second winding of the relay.

The direction of movement of the armature depends upon whether the pilot level is increased or decreased. Thus, if the level should increase by .5 db or more, the armature would be moved to the right contact. This would cause the operation of relay B which in closing connects 60-cycle voltage to the right winding of the "telechron" motor in the regulating amplifier circuit, causing it to rotate in such a direction as to increase the loss of the regulating network. The telechron motor will continue to operate until the control relay is released, correcting at the rate of approximately 1/16 db every 4 seconds. Release is effected by means of the "pulse" relay, the winding of which is connected to a second contact of relay B. The pulse relay is a special mercury delay type which does not operate until four seconds after the path through its winding is closed. When it does operate, a circuit is closed through the second winding of the control relay. This restores its armature to normal, thus releasing relay B and opening the circuit to the telechron motor.

What happens, then, is that when the pilot channel current deviates from normal, the telechron motor operates for four seconds to counteract the effects of this deviation, and then stops. If sufficient correction is not obtained in this time, the operation is repeated. For deviation in the minus direction, the same series of operations occur except that relay A now functions and the telechron motor is rotated in the opposite direction.

The remaining relays shown in Figure 29-7 are provided to take care of sudden large changes in the pilot channel level. Such changes either require manual attention or are of such short duration that a correction would be undesirable. The alarm relay, whose winding is in series with that of the control relay, is adjusted to operate only for large level changes of plus or minus several decibels, as indicated. If such a change occurs, both the alarm and control relays will operate, but the operation of the alarm relay closes a circuit through the winding of the "hold" relay. The latter in turn causes the C relay to operate, which opens the circuit to the telechron motor and thus prevents any change in the regulating network circuit. At the end of four seconds, the operation of the pulse relay restores both the control and alarm relays to normal. The hold relay has a release time of four seconds and if the level change still persists, no correction will be made because relay C remains operated. If the level change was temporary, however, relay C will release after four seconds and normal regulation will be resumed. If a level change of sufficient magnitude to operate the alarm relay persists for a considerable time, the second contact on relay C will cause the "alarm delay" relay to operate after a period of 25 seconds, which will cause the operation of other relays to give visible and audible alarms that the circuit needs manual attention.

29.4 Type-J Carrier Telephone System

The Type-J carrier telephone system provides 12 two-way telephone channels on an open wire pair using a frequency range from about 36 kc to 143 kc. As in the Type-C system, operation is on an equivalent 4-wire basis with West to East transmission employing a frequency band lying between 36 and 84 kc, and East to West a band lying between 92 and 143 kc. Channel carriers are spaced at 4 kc intervals giving an effective channel bandwith of well over 3 kc. Four frequency allocations coded NA, NB, SA and SB are

available for use in limiting crosstalk where more than one J-system is applied to the same line. Since the lowest frequency on the line is 36 kc, it is possible for a pair of wires to carry both a voice-frequency telephone circuit and three Type-C carrier telephone channels in addition to a J-system, thus providing a total of 16 voice circuits.

Figure 29-8 is a block schematic of the essential elements of a West (or South) J-system terminal. It will be noted that the principle of operation does not differ essentially from that employed in Type-C and other carrier systems. Perhaps the most striking new feature is the use of more than one stage of modulation and demodulation. The basic purpose of this is to permit the use of the most desirable group of frequencies for the initial channel modulation, while at the same time applying to the line the desired bands of frequencies. The twelve carrier channel frequencies employed are 64, 68, 72, 76 and so on, up to 108 kc. The separation between carrier frequencies is thus 4000 cycles.

There are a number of reasons for the selection of this group of carrier frequencies for the basic "channel bank". In the first place, it happens that high-grade crystal filters can be economically built for operation in this general range. Also important is the fact that the range is high enough so that the lowest harmonic of the lowest frequency is above the highest frequency of the band. Thus, the second harmonic of 60 kc, which is the lowest frequency in the lower side-band of the 64 kc carrier, is 120 kc, which is well above the top frequency of 108 kc. This obviates the possibility of any harmonics that may be generated in the channel modulators interfering with other channels. Finally, a general design and manufacturing economy is obtained by using this same channel bank for other carrier systems, including Types K, L, and N.

Having modulated the twelve channel carriers with voice frequencies and eliminated everything but the lower side-bands by means of appropriate band filters, the entire group of frequencies—48 kc in width—is translated by an additional modulation process to the band of frequencies that it is desired to transmit over the line. In the case of the Type-J system, the range of this final band depends upon the direction of transmission since transmission in both directions is over the same pair of wires. Thus, for the NA frequency allocation, transmission West to East occupies the band between 36 and 84 kc, and transmission East to West is in the range from 92 to 140 kc.

Due to the fact that the two frequency bands transmitted over the line in the Type-J system both overlap the initial 60 to 108 kc band, a direct translation is not practicable. Instead, it is necessary to make the translation in two modulation stages as indicated in Figure 29-9. It may be observed that for transmission East to West the initial 60-108 kc band is delivered to the first group modulator along with a carrier frequency of 340 kc. The output of the modulator includes the upper 400-448 kc side-band which is selected by the transmitting band filter and passed on, through the intermediate amplifier, to the second group modulator. The carrier frequency applied to this modulator is 308 kc and its output therefore includes a lower side-band of 92-140 kc. The upper side-band and other frequencies above the 92-140 kc band are eliminated by the transmitting low-pass filter, so that the input currents applied to the transmitting amplifier, and thence to the line, are in this desired frequency band. The frequency translations occurring on the receiving side of the circuit are, of course, in the opposite direction and may be followed through in a like manner. For transmission West to East, the processes are exactly similar except that here the carrier frequency applied to the second group modulator in the transmitting circuit (and the first group demodulator in the receiving circuit) is 484 kc instead of 308 kc. The resultant lower side-band of this is 84-36 kc, or, when turned over, the desired 36-84 kc band which is applied to the line for transmission in this direction.

The copper-oxide modulator and demodulator units employed are capable of handling only relatively small amounts of energy, so that the transmission levels at both their inputs and outputs are necessarily rather low. As may be seen by referring to Figure 29-8, the input level to the channel modulator is −13 db and the input levels at the first and second group modulators are considerably lower than this. As the transmitting level applied to the line is about +17 and the receiving level applied to the voice terminal is +4, this means that several amplifiers are required in both the transmitting and receiving legs of the circuit. All of these amplifiers are of the stabilized negative feedback type, capable of giving substantial gains. The demodulator amplifier (single-stage) is adjustable through a plus or minus 5 db range by means of a potentiometer mounted in the voice-
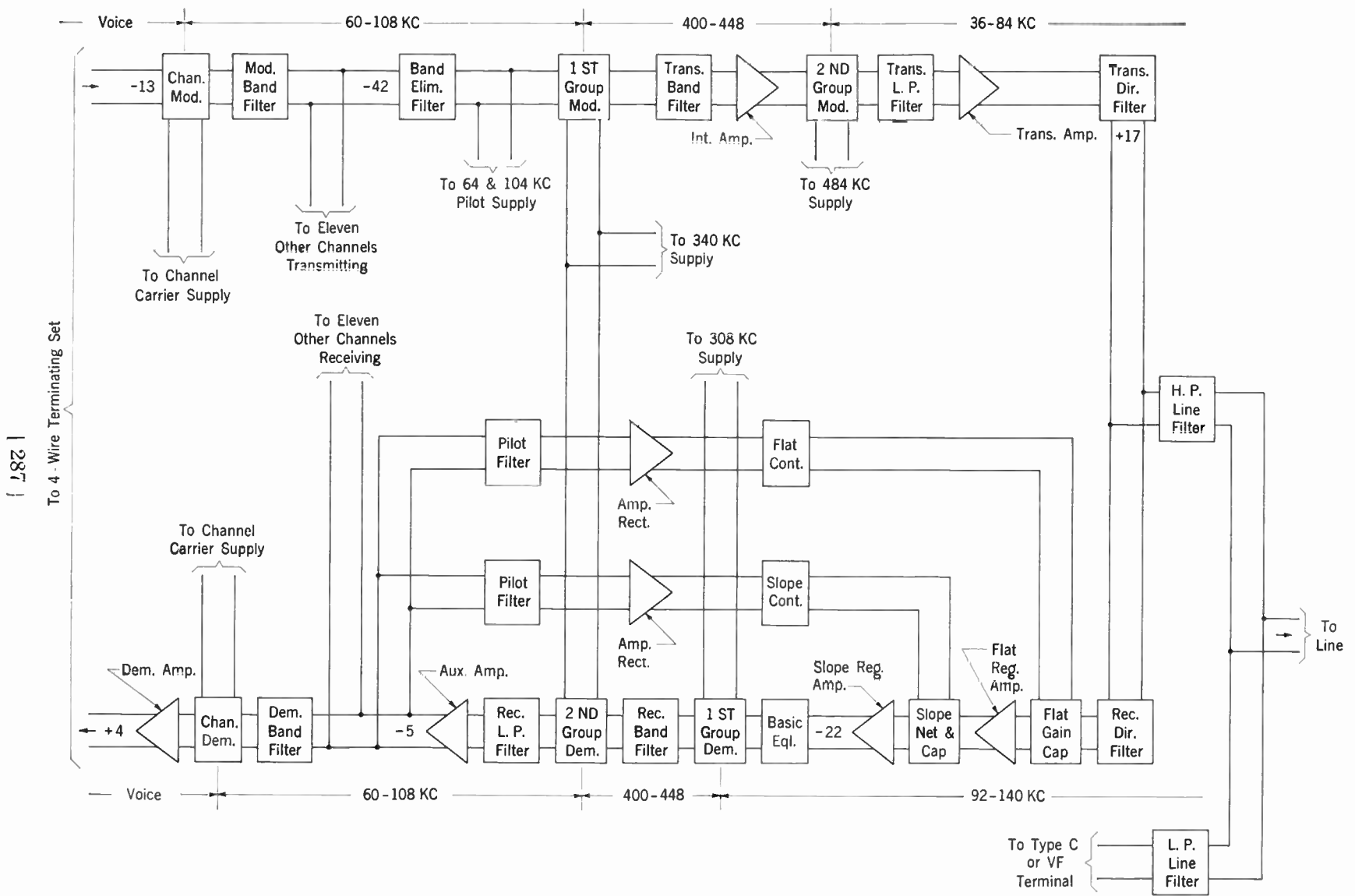
FIG. 29-8 TYPE-J CARRIER TELEPHONE TERMINAL (WEST)

frequency jack panel. This permits convenient manual adjustment of the receiving levels when the circuits are lined up.

Since the line losses at the frequencies employed are relatively high, the Type-J system requires the use of repeaters at considerably closer spacings than does the Type-C. Just what this spacing must be depends upon the weather conditions prevailing in the territory through which the line extends. Sleet, frost, or ice forming on the line wires will greatly increase their attenuation, and the repeater spacing should be close enough so that there will be sufficient margin to take care of the most adverse conditions that may reasonably be anticipated. For most of the lines where these systems are applied, the average spacing is about 50 miles. At the repeater points, the transmission in the two directions is separated by directional filters and each repeater includes two line amplifiers, one "pointed" in each direction.

Automatic regulation to compensate for attenuation variations caused by temperature and weather changes is accomplished in the J-system by means of two pilot channels associated with

each direction of transmission. One pilot channel provides for "flat gain" regulation and the other for "slope" regulation. The flat gain pilot channel frequencies on the line are at 80 kc for W-E transmission, and 92 kc for E-W transmission. The corresponding slope pilots are at 40 and 143 kc respectively. It may be noted in Figure 29-8 that the pilot supply frequencies applied at the input of the first group modulator are indicated at 64 and 104 kc for W-E transmission. The two subsequent frequency translations, however, change these values to 80 and 40 kc respectively when they reach the line. On the receiving side, the pilot channel voltages are picked-off at the output of the auxiliary amplifier, separated by pilot filters, amplified and rectified. The rectified flat gain pilot voltage is fed to a flat control circuit of the same type as was shown in Figure 29-7. This circuit controls the position of a variable flat gain capacitor which in turn determines the gain of the flat regulating amplifier. Similarly, the rectified slope pilot voltage drives a slope control circuit which, through another variable capacitor, controls a slope network and an associated slope regulating amplifier. This slope network,
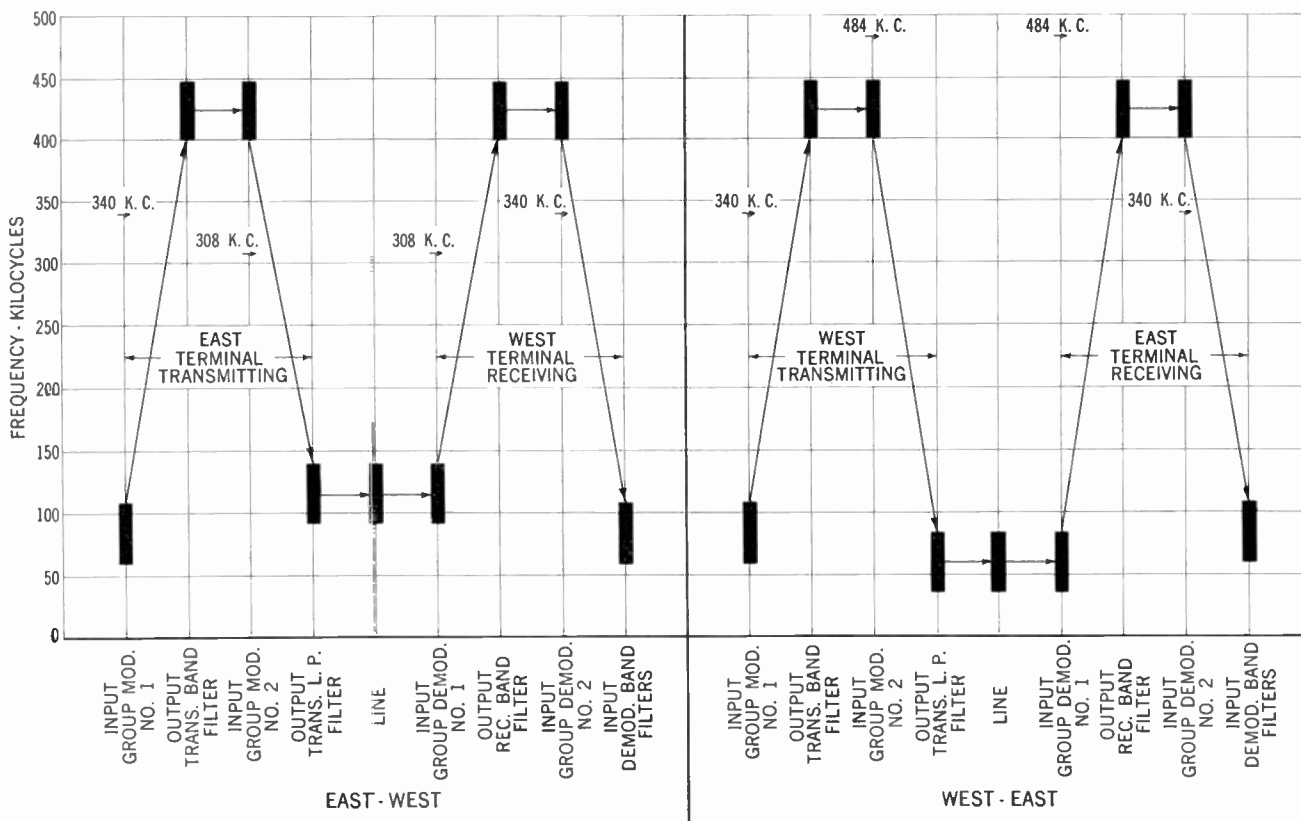
Fig. 29-9 Frequency Translations in Type-J Carrier Systems

[288]
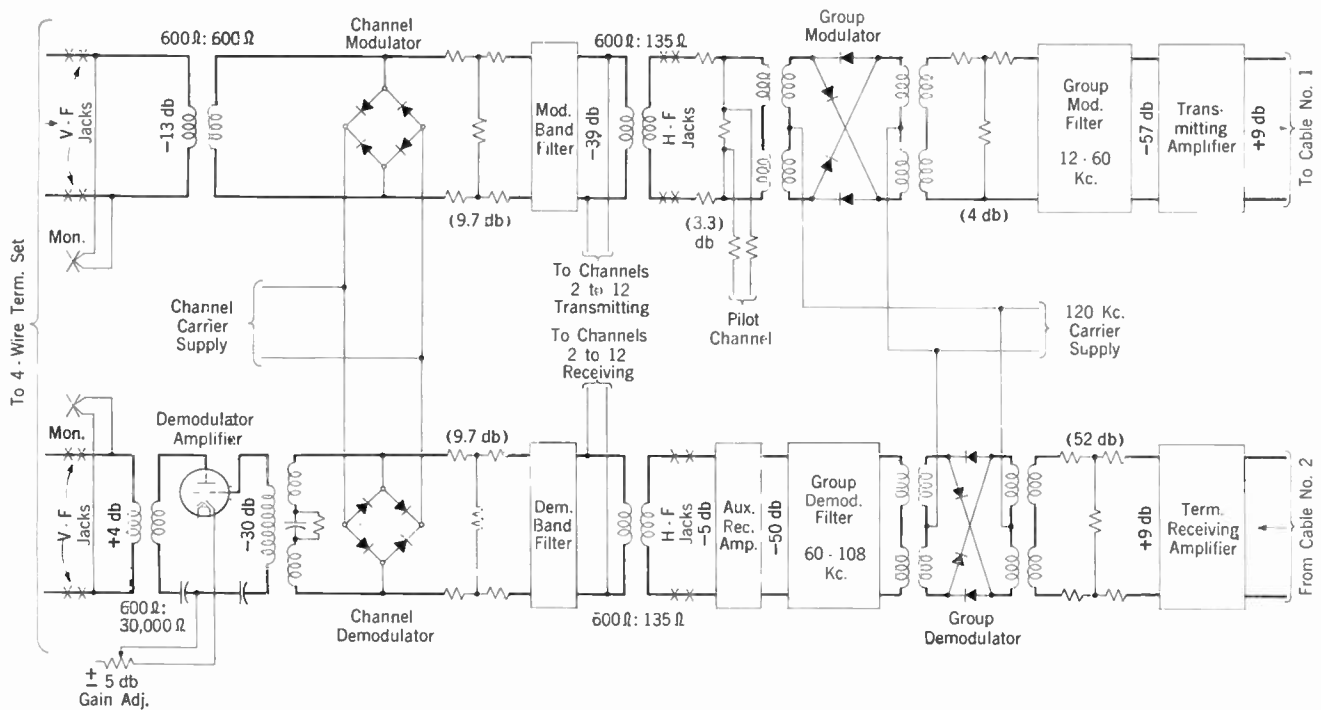
FIG. 29-10 TYPE-K CARRIER TELEPHONE TERMINAL

with its amplifier, is effectively a variable equalizer. It should be noted that the diagram of Figure 29-8 is for a West terminal and the incoming transmission is therefore in an E-W direction. The incoming pilot frequencies from the line are accordingly 92 kc (flat) and 143 kc (slope), which after the two demodulation stages take values of 60 and 111 kc respectively. The regulating circuit arrangement at an East terminal, and the W-E amplifiers of repeaters are somewhat different in detail design but function according to the same general principles.

For discussion purposes, we will consider only the West terminal and the corresponding E-W repeater amplifiers where slope and flat gain regulation are controlled by line pilot frequencies of 143 kc and 92 kc respectively. To understand the overall operation, we may assume weather conditions where the line is changing from dry to wet. The wet line loss will become higher at all frequencies, but the increase in loss will be greater at the higher than at the lower frequencies. The flat gain regulator will adjust automatically to compensate for the increased loss at 92 kc, but there will remain a deficiency of compensation for the higher frequencies. The slope regulating circuit, which is controlled by the high 143 kc pilot frequency, will operate to compensate for this deficiency by increasing the gain at 143 kc, and at

lower frequencies on a gradually decreasing (approximately straight line) basis. Actually the two controls are independent and both actions may take place simultaneously, thus at the same time increasing the total flat gain of the amplifier and changing the slope of the gain-frequency characteristic to compensate for the changed loss-frequency characteristic of the line.

29.5 Type-K Carrier Telephone System

Type-K carrier systems provide 12 two-way telephone channels over two 19-gage non-loaded cable pairs. As shown in Figure 29-10, the terminal channel banks are identical in design with those of the Type-J system, spacing the 12 channels in the 60 to 108 kc band with 4 kc intervals between carriers. For transmission over the line, however, the techniques employed in the two systems are necessarily quite different. Because the attenuation of non-loaded cable conductors is high and of course increases with frequency, it is desirable to keep the maximum line frequency at the lowest practicable value. Accordingly, the band of frequencies selected for transmission on the cable line is that between 12 and 60 kc, which occupies the comparatively straight-line portion of the attenuation-frequency curve just above the knee of the curve. By using pairs in separate

[289]
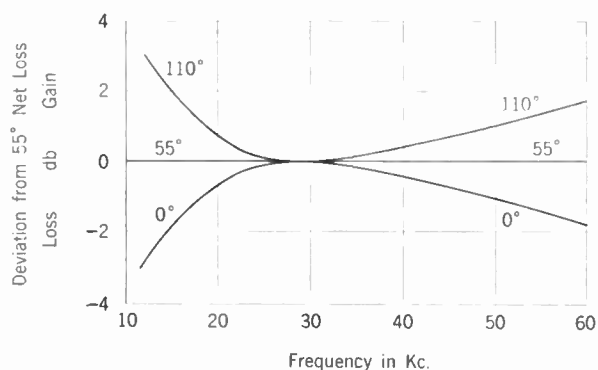
FIG. 29-11 TWIST EFFECT IN 100-MILE AERIAL CABLE CIRCUIT

cables for transmission in the two directions, only one 48 kc band is required. On this basis, the line losses are such as to require the insertion of high-gain repeaters at intervals averaging about 17 miles.

At the system terminals, a single group modulator is used to translate the initial 60 to 108 kc band to the 12 to 60 kc band applied to the line. This is the inverted lower side-band of a 120 kc carrier supplied to the group modulator. After passing through the transmitting amplifier, the output level applied to the line is +9 db. In the same way on the receiving side, the group demodulator translates the incoming 12-60 kc band back to the 60-108 kc band, which again is the inverted lower side-band of the 120 kc carrier.

Three basic types of amplifiers are used in K-systems for counteracting normal line attenuation, and for gain regulation to compensate for variations in line attenuation. Such variations in cable facilities are due entirely to temperature changes, since the insulation of the cable conductors is not affected by weather conditions. The three types of amplifiers are known as terminal transmitting amplifiers, line amplifiers and twist amplifiers. The latter two types of amplifiers furnish automatic regulation as well as flat gain. The terminal receiving amplifier is a twist amplifier.

Variations in attenuation of cable facilities due to temperature changes may be considered as being composed of two components. The first of these is a "flat loss" variation, which is independent of frequency. In addition, changing temperatures cause attenuation variations which are not exactly the same at all frequencies in the transmitted range. In other words, the shape of the attenuation-frequency characteristic of cable facilities may be different at different temperatures. This effect is known as "twist". The curves of

Figure 29-11 indicate the general magnitude of the twist effect for a 100-mile aerial cable circuit over the Type-K frequency range, in terms of deviations from the loss at 55°F. Compared in magnitude to the flat loss variation resulting from temperature changes, the twist variations may appear insignificant. Nevertheless, they are too large to be allowed to accumulate over a long circuit. Automatic regulation is accordingly required in practice at intervals of about 100 miles in aerial cable, and 200 miles in underground cable, to take care of the twist effect. Flat gain regulation, in contrast, is required at every repeater point to take care of the flat loss variation of the adjacent 17-mile cable section.

In the original design of K-systems, both flat and twist regulation are provided by means of pilot wires extending through one repeater section for flat gain regulation, and through a twist section (several repeater sections) for twist regulation. The loss variations in these pilot wires control mechanical devices associated with the regulating amplifiers, in a manner similar in principle to the voice-frequency regulating circuit discussed in Article 28.6. In more recent designs, regulation is secured through pilot channels employing frequencies of 12, 28, 56, and 60 kc on the line. The 60 kc channel controls the flat gain of the line amplifiers while the other three channels control the twist amplifiers. Changes in the regulating amplifier gains are effected in all cases through the use in the feedback circuits of thermistors whose resistance is varied by the pilot channel currents.

The thermistor is a circuit element made of a number of semi-conducting metallic oxides, which have a large negative temperature coefficient. In contrast to the usual conducting metals, which have small positive temperature coefficients, the resistance of a typical thermistor decreases in the order of 2.5% per degree Fahrenheit so that its resistance is halved for about every 27°F increase in its temperature. Several types of thermistors are used in K-carrier systems. Physically, these take the form of a tiny "bead" of semi-conducting material not much larger than a pin-head, or of a "disc" of like material about the size of a 5-cent coin. The disc type is used as a thermostatic control device whose resistance is determined by the ambient temperature. The bead thermistor may be either directly or indirectly heated, but in either case is surrounded by a tiny coil of nichrome wire, which is insulated from the bead

itself and may be used to vary the temperature of the bead and thus change its resistance. Where slow action of the thermistor is desired, the bead is enclosed in glass to increase the thermal capacity, and a thin film of gold is deposited on the outside of the glass to reduce heat loss by radiation.

In the line amplifiers of K-systems, a directly heated thermistor bead is inserted in series with the amplifier negative feedback path as indicated in Figure 29-12. A portion of the amplifier output current, whose total value is controlled by the 60 kc pilot channel, is fed back through this thermistor into the amplifier input. If the amplifier output tends to decrease below its normal value as a result of increased attenuation in the preceding line section, the current through the thermistor will tend to decrease accordingly. This causes the thermistor resistance to increase, which reduces the total current fed back into the amplifier input and thus increases the amplifier gain in amount sufficient to compensate for the increased line attenuation.

Figure 29-13 shows how the thermistor resistance varies with the current flowing through it. It will be noted that the normal operating range of this thermistor is between 1100 and 20,000 ohms, which corresponds to current values of 1.7 and .15 ma. In order that the thermistor operation be uniform and within this operating range, the thermistor bead is maintained at a constant
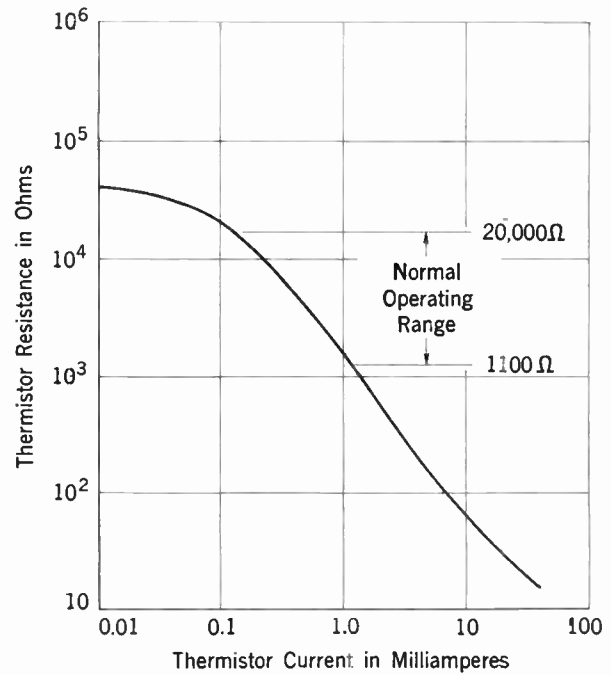
FIG. 29-13 CHARACTERISTIC OF FLAT GAIN REGULATING THERMISTOR IN LINE AMPLIFIER

reference temperature of 160°F by means of an auxiliary current through its heater winding. This current is supplied from an 8-volt filament supply tap, as indicated by Figure 29-12. To avoid changes in the bead's temperature as a result of ambient temperature changes, a disc-type thermistor is bridged across the heater winding sup-
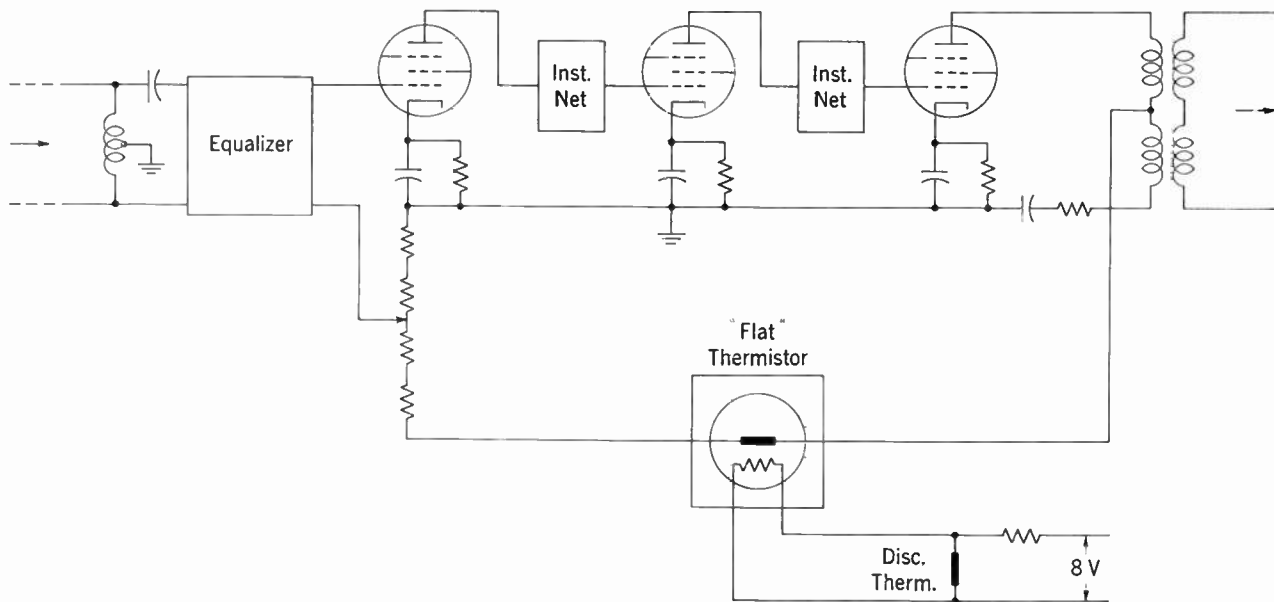
FIG. 29-12 TYPE-K LINE AMPLIFIER-REGULATING

ply. The disc is mounted close to and in the same container as the bead, but because of its size, its temperature is controlled by the surrounding ambient temperature and not affected by the temperature of the bead itself. Accordingly, if the ambient temperature should increase, the resistance of the disc will decrease thus reducing the heater current by its increased shunting effect. The thermistor circuit is arranged to be relatively slow acting (10 to 15 minutes) to correspond to the normally slow variation of the line attenuation.

The design objective of the regulating line amplifier is to maintain practically constant total output power at each amplifier, regardless of possible substantial variations in input power. To accomplish this objective, it is necessary that the same constant power value be supplied to the line by the terminal transmitting amplifier. This total power output consists of the power of the 60 kc pilot channel plus the combined power of the signal channels. Its value in the K-system is +15 dbm (15 decibels above 1 milliwatt). The transmitting amplifier accomplishes the double purpose of raising the signal currents to a proper level for transmission over the line, and of generating the 60 kc pilot channel power. The arrangement is such that when no signals are being transmitted, the 60 kc output power alone will total +15 dbm. When signal currents are present, the amplitude of the 60 kc oscillations automatically decreases to whatever extent is necessary to still maintain the total output power at +15 dbm. This means that the output of the transmitting
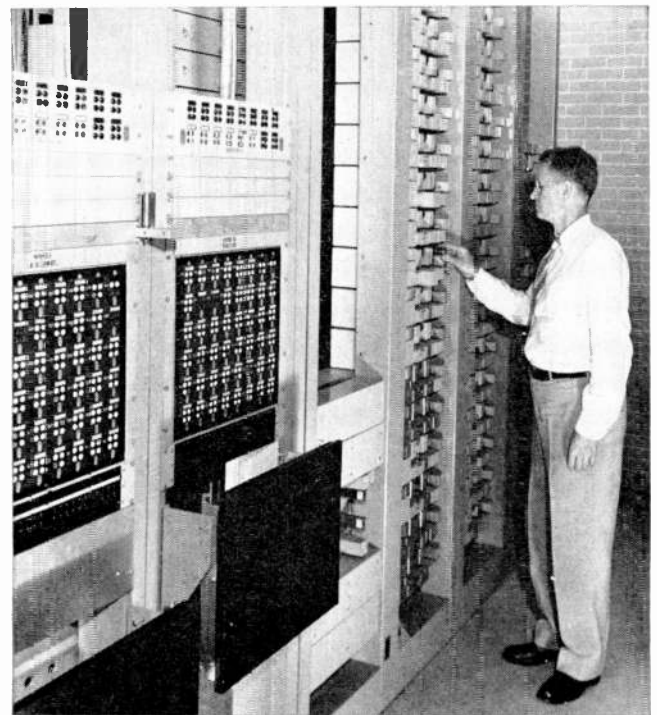
TYPE-K CARRIER REPEATERS AT AUXILIARY STATION

amplifier, and the gains of the succeeding line amplifiers, do not depend upon whether any or all of the 12 signal channels are in use.

The design of this transmitting amplifier-oscillator is shown in simplified form in Figure 29-14. It will be noted that there are two distinct feedback paths. One of these is a standard negative feedback which includes a gain adjusting pad by which the amplifier gain may be set at its desired
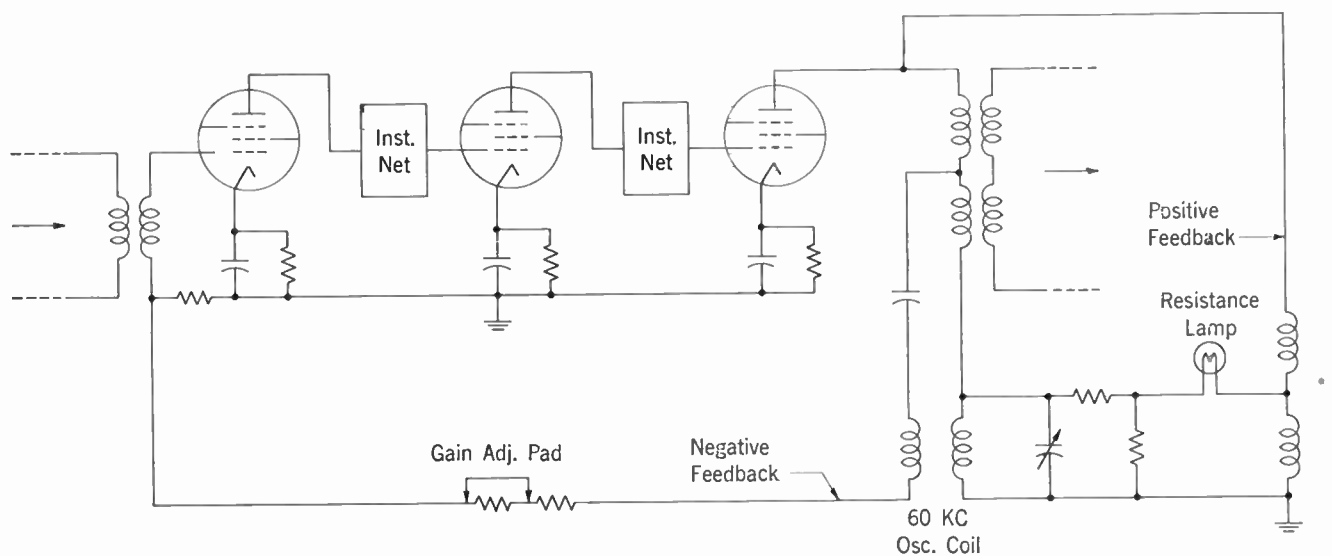
FIG. 29-14 TYPE-K TRANSMITTING AMPLIFIER
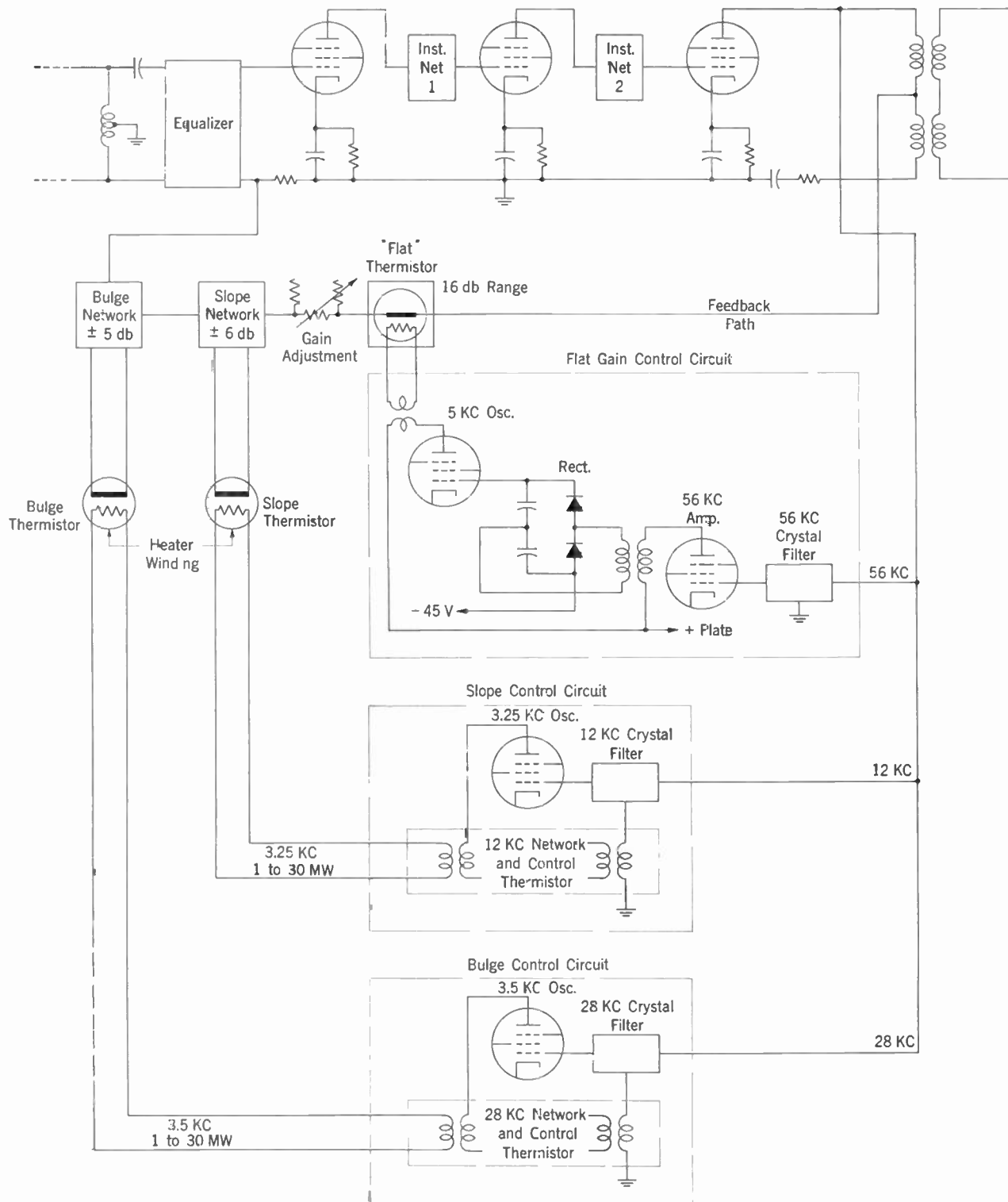
FIG. 29-15 SCHEMATIC OF TWIST AMPLIFIER

value. The other feedback path, which includes a resistance lamp having a negative temperature coefficient and an oscillator coil, is positive. Oscillations set up in the 60 kc tuned tank circuit will in-

crease in magnitude until the current flowing in the resistance lamp reduces its resistance to a point where the loss in the positive feedback path is just equal to the amplifier gain. Currents flow-

[293]

CHAPTER 30

CARRIER SYSTEMS—(*Continued*)

30.1 Type-L Carrier Systems

Type-L carrier telephone systems are designed for application to coaxial conductors. The telephone terminal equipment may also be used on microwave radio systems or other mediums capable of handling an extremely wide range of frequencies. The original system, coded L1, was designed to provide 480 two-way telephone channels over a pair of coaxial conductors, using the frequency range between 68 and 2044 kc. Later development added 120 more channels for relatively short distance transmission, thus extending the frequency range utilized up to 2788 kc. The 600 voice channels, each 4 kc wide, are placed in their line frequency allocations by means of three modulation steps in the manner indicated in Figure 30-1. The first modulation step places 12 voice channels in the 60 to 108 kc range to form a channel bank identical with the channel banks used in the J and K systems, as previously discussed. In a second step of modulation, five channel banks are translated to the frequency band between 312 and 552 kc. This constitutes a **basic supergroup** of 60 voice channels. The final modulation step translates the supergroups to appropriate line frequency positions as shown in Figure 30-1, which also indicates the carrier frequencies used in the group and supergroup modulators. It may be noted that supergroup No. 2 does not require modulation but is placed directly on the line in the 312-552 kc position. In practice, of course, any number of supergroups up to the total of 10 may be used, depending upon the total circuit requirements.

Figure 30-2 shows diagrammatically the essential elements of the L1 terminal. The group and supergroup modulators and demodulators employ copper-oxide disc varistors in a bridge arrangement. Intermediate amplifiers are included as required to compensate for the losses caused by the filters and hybrid coils, which form a necessary part of the total circuit. A group pilot of 92 kc is applied at the input of the modulator for each group, as indicated in Figure 30-2. On the receiving end, this pilot frequency is picked off at the output of the auxiliary amplifier following the

group demodulator to provide a rapid means of trouble location. It may also be used as a means of system lineup and as a convenient check on transmission.
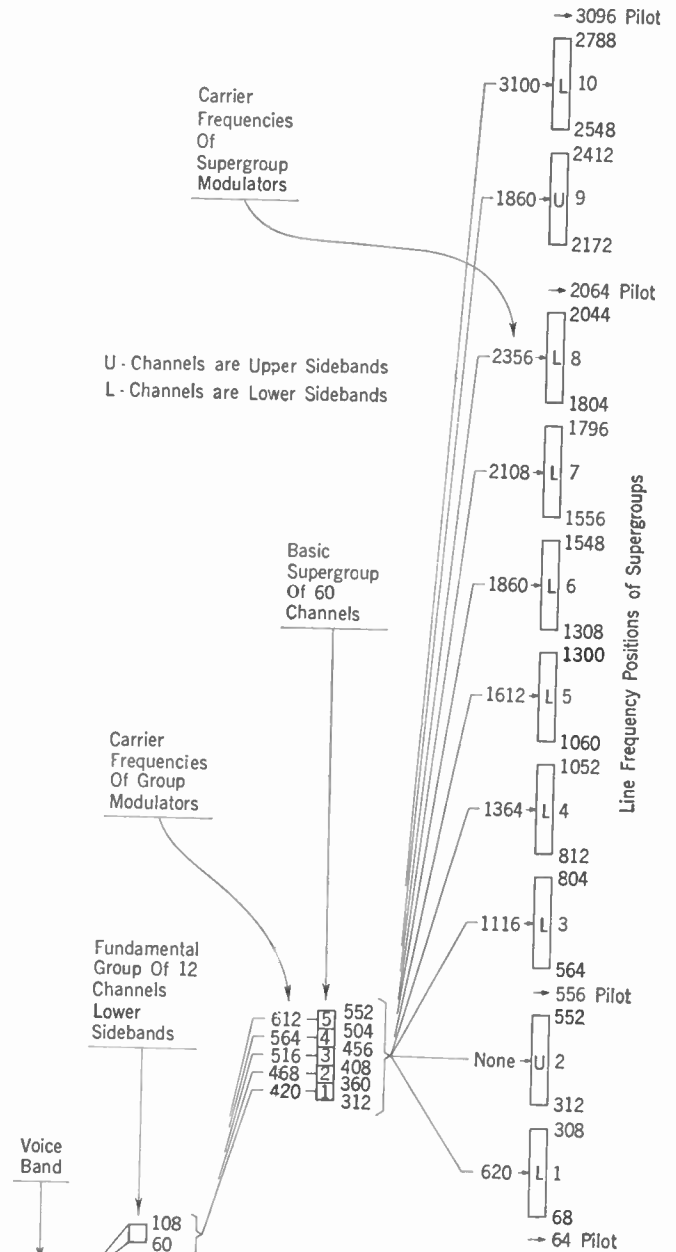
Fig. 30-1 Frequency Translations of L1 System
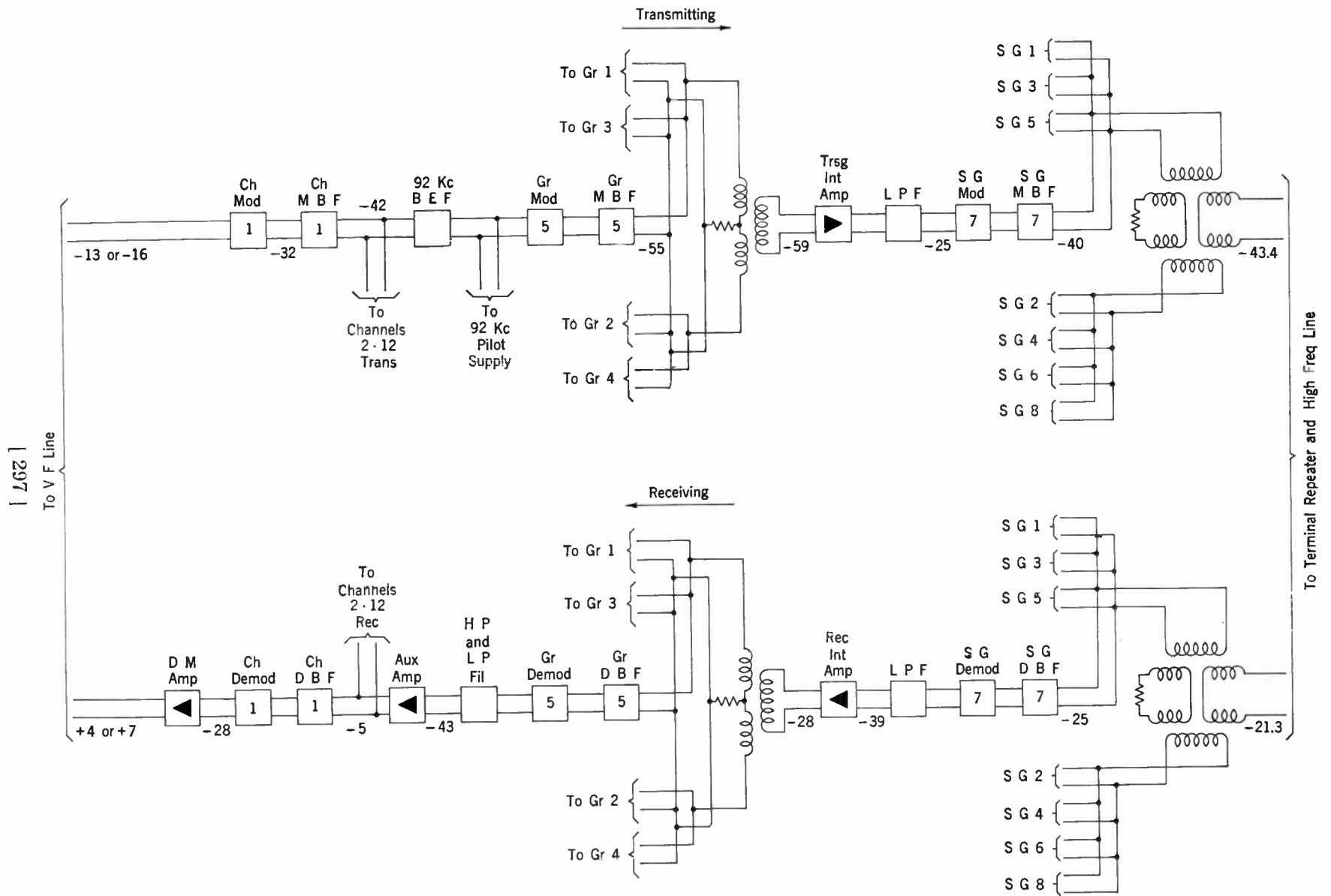
FIG. 30-2 TYPE-L1 CARRIER TERMINAL

Transmission over the high frequency line of the wide band of frequencies to which the 600 voice channels of the L1 system are translated requires the use of many amplifiers and associated regulating and equalizing equipment. The general arrangement for transmission in one direction is shown in block schematic by Figure 30-3. The line section illustrated is designated as a "switching section". Circuit dropping or branching cannot occur within a switching section. These sections, therefore, may extend from one terminal to a distant terminal; or to an intermediate point which may or may not be arranged for terminating some part of the carrier channels, or for connecting some or all of them to other transmission paths. In the case of .375 coaxial, amplifiers must be located along the line at intervals of approximately 8 miles. Most of these are "auxiliary repeaters" which consist only of amplifiers, basic equalizers, and associated regulating equipment.

At intervals usually in the order of about 100 miles but in no case greater than 165 miles, "main

repeaters" which include adjustable equalizers as well as amplifying, regulating and power supply equipment, are provided. These main repeaters may be of the switching or non-switching type depending on maintenance and overall circuit layout requirements.

Sixty-cycle a-c power for the operation of auxiliary repeaters is fed from terminal and main repeater points over a series loop made up of the two center conductors of the pair of coaxials used for the two directions of transmission. The 60-cycle currents are separated from the high frequency transmission currents by means of power separation filters. Since as many as eleven auxiliary repeaters and a terminal or half a main repeater must sometimes be so supplied in series, the voltage applied at the source may be quite high (up to 2000 or more volts) in order to overcome the resistance drops along the line, and the drops in the heater elements of the vacuum tubes of the repeaters served.

As indicated in Figure 30-3, the output of the
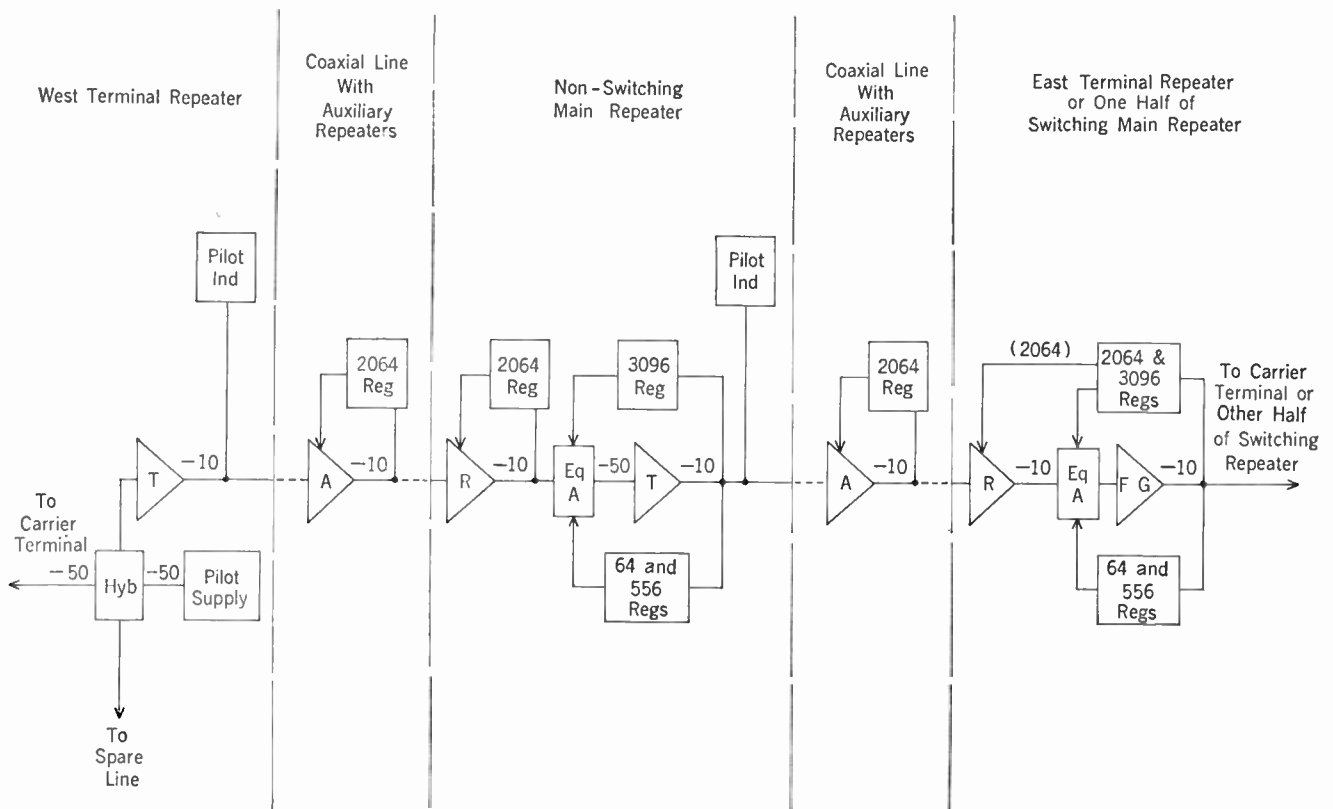
FIG. 30-3 BLOCK SCHEMATIC OF L1 CARRIER SWITCHING SECTION

transmitting terminal is supplied to the terminal repeater at a level of −50 db. It is transmitted to both the working line and a spare line by means of a hybrid coil. The spare line circuit is not shown in the diagram but is identical with that of the working line. At the transmitting end of a repeater section, a transmitting amplifier of the type described in Article 25.2 raises the level to −10 db and applies it to the coaxial line. Auxiliary amplifiers along the line counteract the losses in each preceding 8-mile section of line, bringing the levels up again to −10 db at their outputs. At the receiving end of a section, switching main and terminal repeaters include a basic equalizer and receiving amplifier, which serves the same purpose as the auxiliary amplifiers, adjustable equalizers and one or more flat gain amplifiers. Non-switching main repeaters are similar except that flat gain amplifiers are not normally required.

At the transmitting hybrid coil, pilot voltages are supplied at the four frequencies of 64, 556, 2064 and 3096 kc. The 2064 kc pilot is used to vary the gains of the auxiliary and receiving amplifiers to compensate for line attenuation changes caused by temperature variations. The regulators are of the dynamic type, acting to hold the 2064 kc pilot power at the output of each regulated amplifier at a practically constant value. Their design is similar to that of the regulator shown in Figure 29-15. The 2064 kc pilot is picked off at the output of the amplifier by a selective filter, amplified and rectified (in this case by means of a vacuum tube). The rectifier output is applied to the grid of a 2000-cycle vacuum tube oscillator whose output in turn drives the heater element of an indirectly heated thermistor inserted in the feedback circuit of the amplifier. As in the case of regulating circuits previously discussed, the design is such that a small change in the pilot input causes a relatively large change in the thermistor heating current and, consequently, in the amplifier gain.

Where coaxials are in underground or buried cable, it is the general practice to use automatic regulation only at every other auxiliary repeater. Alternate repeaters are then arranged for manual regulation. Such regulation is also effected by means of the thermistor in the amplifier feedback, which is indirectly heated by current supplied through a manually adjusted potentiometer. In this case, a disc-type thermistor is associated with the gain control thermistor to compensate for the

effect of ambient temperature changes.

Because of the very wide band of frequencies employed, it is evident that equalization in Type-L systems presents a larger problem than in other carrier systems. Major equalization is effected, as in all other transmission systems, by means of fixed equalizers which are inserted at appropriate points, although not shown in Figure 30-3. To take care of variations in line and apparatus transmission characteristics that are functions of temperature or frequency, special equalizers, designated equalizer A in Figure 30-3, are employed at main and terminal repeater points, and at "equalizing auxiliary stations". These latter are similar to the non-switching main station shown in Figure 30-3 except that power supply arrangements are omitted. The equalizers are actually a series of networks, each designed to modify the shape of the transmission characteristic over a partial range of the total frequency band. Some of them are manually adjustable and others are adjusted automatically under the control of the 64, 556 and 3096 kc pilots. In the latter cases, dynamic regulating circuits similar to that used by the 2064 kc pilot, control the resistance of thermistors which form a part of the equalizing networks. Pilot regulators are omitted at equalizing auxiliary stations and manual adjustment of the A equalizer at these points is necessary.

The same kinds of coaxial conductors that are used for Type L1 carrier transmission are also a satisfactory transmission medium for Type L3 carrier, which operates over a much broader frequency band than L1. The L3 system is capable of providing as many as 1860 two-way telephone channels in the frequency range between 312 and 8284 kc. The general method of accomplishing this is indicated in Figure 30-4. Ten 60 channel supergroups, formed in the same manner as in the L1 system, are modulated with appropriate carriers to form a **master group** of 600 voice channels. The first such master group is placed in the frequency range between 564 and 3084 kc. A second master group of 600 channels, formed in the same way, is placed between 3164 and 5684 kc, and a third group between 5764 and 8284 kc. In addition, a single supergroup may be transmitted below master group 1 in the basic supergroup range between 312 and 552 kc.

For application on .375 inch coaxials, the L3 system requires repeater spacing at 4-mile intervals instead of the 8-mile intervals of the L1 system in order to maintain the high-frequency

energy along the line at workable levels. The amplifier design must also be different because of the extremely broad frequency band transmitted. Each amplifier employs three vacuum tube tetrodes and two triodes, all having very high transconductance. These are connected in two separate units (input and output), each being a two-stage amplifier with its own feedback circuit. The two units are connected in tandem through a regulating network to form the complete amplifier.

Power for the operation of the many auxiliary repeaters in each switching section is supplied from main repeater stations in the same way as in the L1 system. Since the number of intermediate repeaters is doubled, however, the maximum

voltage between the two coaxial center conductors may be considerably higher than in the L1 system.

Regulation in L3 systems employs six pilot frequencies at 308, 556, 2064, 3096, 7266 and 8320 kc. The 7266 kc pilot controls amplifier gains compensating for line attenuation changes due to temperature variations. The other five pilots control adjustable equalizers at main and terminal repeater points. Dynamic regulation under the control of the 7266 kc pilot is used, in general, at every other auxiliary repeater. The alternate repeaters employ a method known as **thermometer control** regulation, in which a thermistor, mounted in a waterproof container, is buried near the underground coaxial cable to sense variations in ground
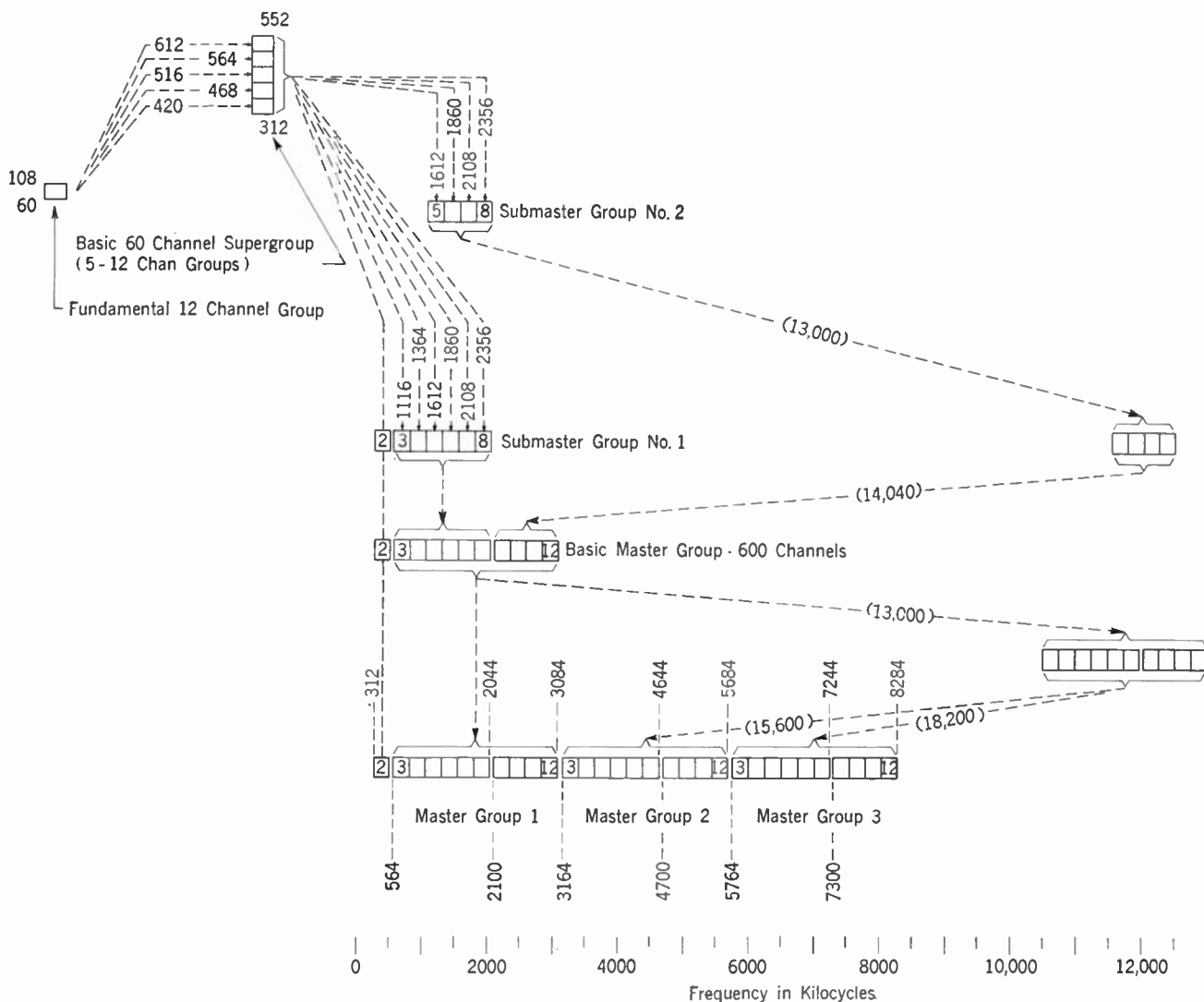
FIG. 30-4 FREQUENCY TRANSLATIONS OF L3 SYSTEM

[300]

temperature. This thermistor in turn controls a thermistor inserted in the amplifier feedback circuit.

30.2 Television Transmission

Type-L carrier system facilities are used for the transmission of television signals as well as for multiple channel voice transmission. Transmission of a television signal necessarily requires the employment of a very wide band of frequencies. This results from the fact that, television depends upon the repetitive detailed scanning of a scene at extremely rapid intervals. Standard practice in the United States for black and white television calls for 525 horizontal lines for each complete scanning of the scene and for 30 complete scans per second, with the reproduced image having a width to height ratio of 4 to 3. In practice, a single complete scan or "frame" is accomplished in two steps. In the first step, the scene is scanned over the odd-numbered 262-$\frac{1}{2}$ lines to form one "field"; and in the second step, it is again scanned over the even 262-$\frac{1}{2}$ lines. This procedure, known as interlaced scanning, affects the eye of the viewer of the image as if the total scene were being reproduced 60 times per second instead of 30, and thus minimizes "flicker".

The scanning sequence is shown in Figure 30-5. For each line, the electron beam in the television camera and in the cathode-ray receiving tube moves horizontally across the image. At the same time it moves vertically downward a distance corresponding to two lines, under the control of the sweep circuit voltages applied to the deflecting
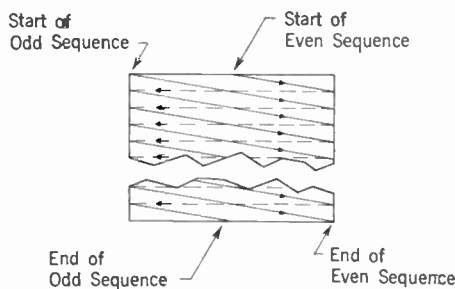
FIG. 30-5 TV SCANNING SEQUENCE

plates or coils of the tubes. The scanning beam is blanked out at the completion of each horizontal line and returned quickly to the starting point of the next line, as indicated by the dotted lines in the Figure. The process is repeated until the bottom of the image is reached. The beam is then blanked out for a longer interval while it is re-

turned to the top of the image for the start of the next scanning sequence. The duration of each scanning line is 54 microseconds and 9.5 microseconds are allowed for the horizontal retrace of the beam. The image is scanned at the rate of 15,750 lines per second.

To maintain the exact synchronization between the camera and the receiver that is obviously necessary, synchronizing pulses generated at the image pickup point are applied to the camera tube and transmitted to the receiver along with the image signals. The synchronizing pulses are superimposed on the signal blanking pulses in such a way that they can be "clipped" from the image signal and applied to the saw-tooth generators which control the deflections of the scanning beam. As previously noted, the horizontal synchronizing pulses must recur at the rate of 15,750 per second and the vertical pulses, which return the beam from the bottom to the top of the image, must recur at the rate of 60 times per second.

Figure 30-6 (A) indicates graphically the form
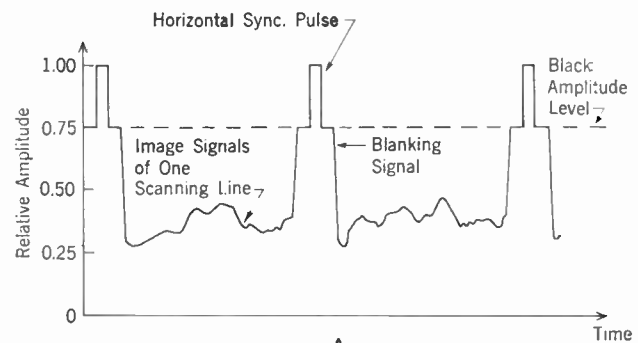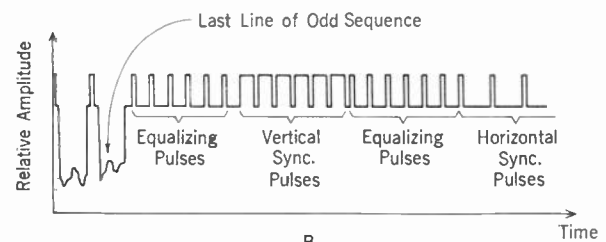
FIG. 30-6 TV SIGNAL FORM

of the TV signal at the receiver for two scanning lines covering a total time of 127.0 microseconds. The image signal, which is applied to the control electrode (grid) of the picture tube, may vary between zero amplitude for "white" and an amplitude which effectively blocks the electron beam to produce "black" in the image. The synchro-

nizing signals, it may be noted, rise above the black level to a region sometimes called "blacker than black". Figure 30-6 (B) illustrates the form of the longer vertical synchronizing pulse, which extends over a period of 190.5 microseconds. Vertical and horizontal synchronizing pulses are separated for application to their proper respective deflecting coils by means of a simple *RC* timing circuit which recognizes the large difference in their time duration. The vertical pulse is "serrated" as shown so that the horizontal pulses will continue during the vertical deflection period to avoid the possibility of their falling out of step. A series of "equalizing" pulses is included before and after the vertical synchronizing pulse to take care of the time factors introduced by the fact that the first scanning field is completed in the middle of a line, and the second at the end of a line.

Considering the transmission of the total television signal, it is evident that the indispensable synchronizing pulses alone make the signal rather complex. The part of the signal carrying the image must be much more complex if satisfactory image detail (resolution) is to be obtained. Thus if a scene is to be analyzed as the horizontal beam crosses it in the same detail as is provided by the 525 line dissection of the image vertically, the signal might take 4/3 x 525 or 700 different values for each horizontal trace. This would correspond to a variation at the rate of 350 cycles per line which would mean 350 x 525 x 30 or approximately 5-1/2 million cycles per second. Furthermore, if the scene being televised was one in which there were many transitions between black and white, such as a black and white checkerboard pattern, the image signal would tend to take the form of a square wave. Accurate transmission in such a case would theoretically involve frequencies extending toward infinity. Actually, practical experience indicates that entirely satisfactory resolution for black and white images is obtained from a video signal including frequencies up to a maximum of about 3 million cycles, although the standard broadcast TV signal is normally considered as 4 million cycles in width. In any event, it is to be noted that the lower frequencies are indispensable. Included here are the vital synchronizing pulses as well as the major values in the image structure. The higher frequencies become increasingly less important as they approach values which tend to enhance the detailed accuracy of the picture beyond the practical limit of perception of the normal eye. As might be expected also, the major energy content of the signal tends to be concentrated in the lower frequencies.

For transmission over the Type-L1 carrier system, the frequency range on the line between about 200 and 3100 kc is employed. The lower frequency is limited by equalization difficulties and the upper by the characteristics of the line repeaters. Since the standard video signal begins at about 30 cycles and may be considered as extending upward, in this case, to about 2800 kc, it is necessary to translate it by modulation procedures to place it in the proper position for transmission over the line. This is accomplished by two stages of modulation, as indicated in Figure 30-7. The carrier frequency of the first modulation stage is 7944.72 kc. A band-pass filter permits the pas-
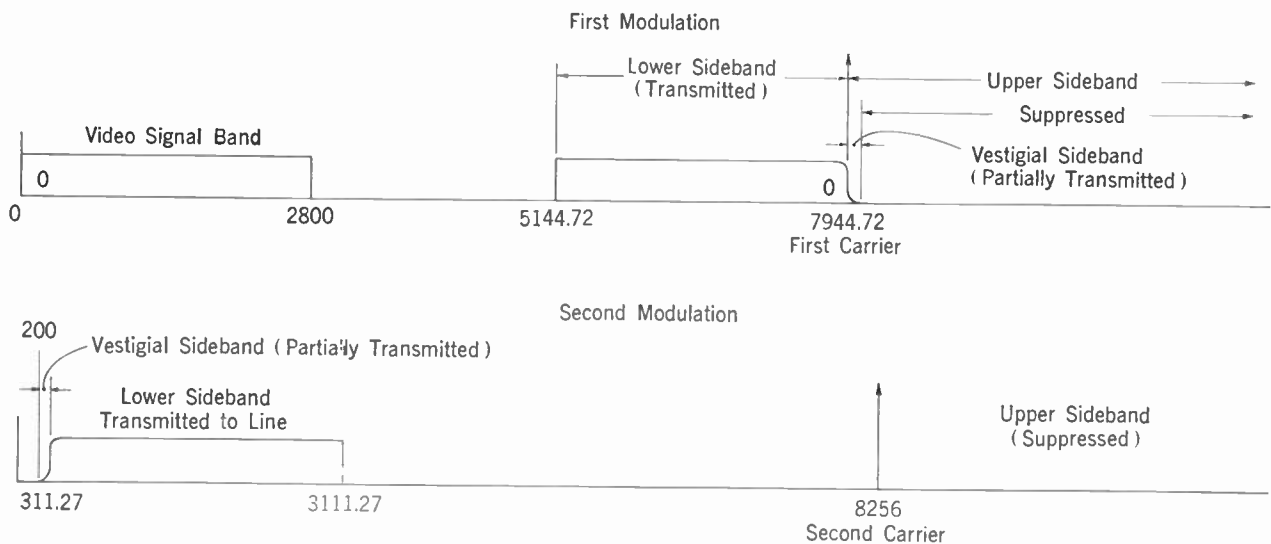
Fig. 30-7 Video Signal Frequency Translations
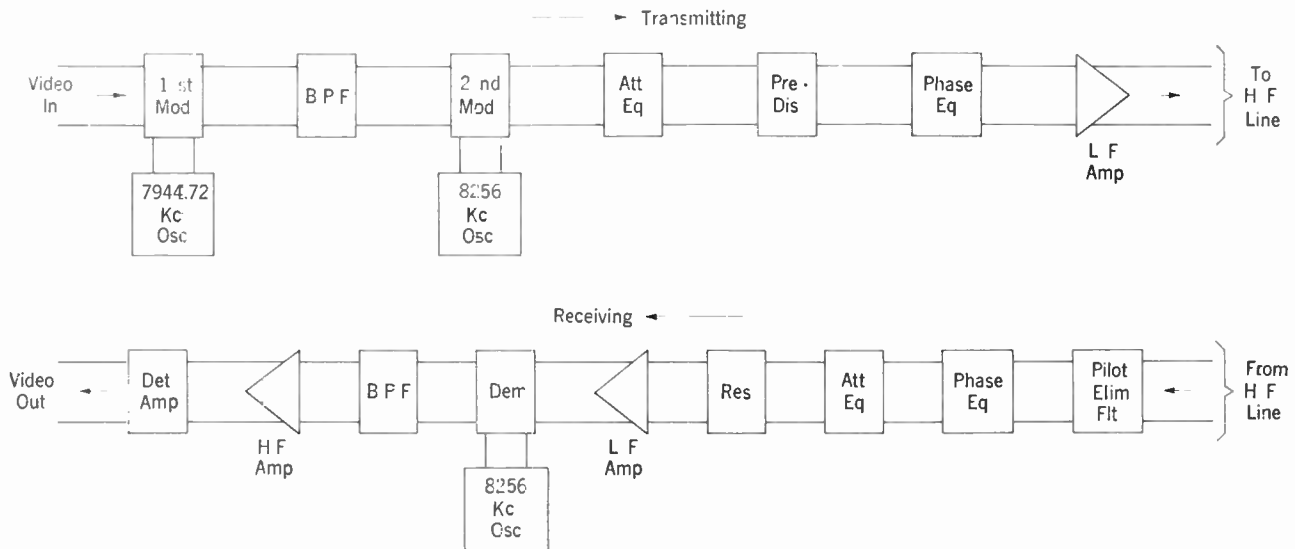
FIG. 30-8 BLOCK SCHEMATIC OF TELEVISION TERMINAL

sage of the lower side-band, extending from about 5100 kc up to the carrier frequency, plus a small part of the upper side-band, extending from the carrier frequency up to about 8100 kc. This latter is known as a "vestigial sideband" and is included in the passed band to insure first, that there is no clipping of the lower side-band; and second, to reinforce the lower frequencies of the signal which, as has already been noted, are of vital importance. The second modulation stage employs a carrier of 8256 kc to translate the foregoing main and vestigial sidebands to the range between about 200 and 3100 kc, with the carrier frequency now appearing at 311.27 kc.

A block schematic of the transmitting and receiving television terminals is given in Figure 30-8. The first modulator employs a balanced vacuum tube circuit to which the 7944.72 kc carrier is supplied by a crystal oscillator. The modulator is preceded by a two-stage balanced vacuum tube video amplifier, the second stage of which is a cathode-follower. The second modulator employs balanced vacuum tubes with input and output transformers. To maintain transmission flat and without delay distortion over the wide frequency range, both attenuation and phase equalizing networks are required at the transmitting and receiving terminals. The phase equalizing networks are needed principally to compensate for delay contributed by the band-pass filters.

The transmitting terminal also includes a special network known as a **pre-distorter**. As shown in Figure 30-9 (A) this inserts a loss in the im-

mediate neighborhood of the carrier frequency (311.27 kc) having a maximum value of about 18 db. The purpose is to reduce the relatively large energy values of the low frequencies and thus to
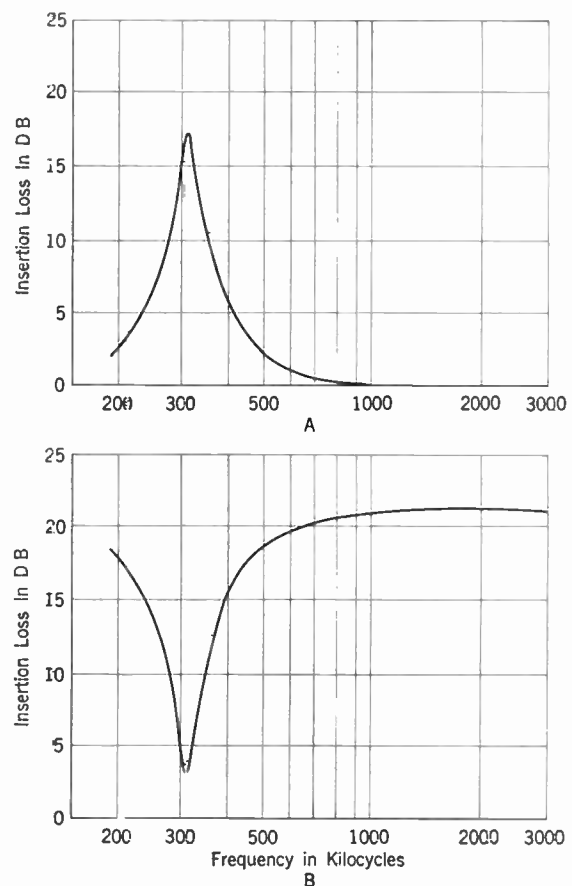
FIG. 30-9 CHARACTERISTICS OF PREDISTORTER AND RESTORER

[303]

distribute the energy of the television signal applied to the line more evenly over the entire frequency range. At the receiving terminal, the signal is restored to its original form by a **restorer** network whose complementary characteristic is shown in Figure 30-9 (B). A three-stage line frequency amplifier raises the signal level to the proper value for application to the high-frequency line.

The first apparatus unit in the receiving terminal is a pilot elimination filter which removes the 556, 2064 and 3096 kc pilot frequencies used for regulation in the Type-L1 carrier system. These would naturally cause distortion of the received image if permitted to pass through to the demodulator. A demodulator, employing a carrier frequency of 8256 kc translates the incoming signal to the frequency range between about 5100 and 8100 kc. Its output, after passing through a band-pass filter, is amplified by a three-stage high-frequency amplifier having a nominal gain of about 31 db and transmitted to a detector-ampli-

THERMISTOR ASSEMBLY USED FOR GAIN REGULATION OF REPEATER AND GROUP AMPLIFIER
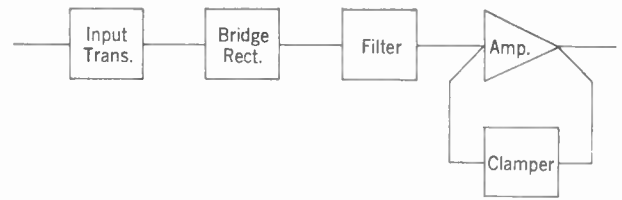
FIG. 30-10 BLOCK SCHEMATIC OF DETECTOR-AMPLIFIER

fier circuit which reproduces the original video signal. The employment of linear detection instead of a second stage of demodulation eliminates the difficult problem of supplying a local carrier frequency that would have the same frequency and the same phase as the carrier at the transmitting terminal.

As indicated by the block schematic of Figure 30-10, the detector-amplifier includes an input transformer, a germanium varistor bridge detector, a filter to eliminate the second harmonic of the transmitted carrier, and a clamper-amplifier. The amplifier proper is of three stages, with the third stage including four triodes connected for parallel operation. The clamper circuit, which is bridged around the amplifier, is a "d-c restorer". The TV signal at the output of the camera consists of a varying direct current which may be considered as an alternating current superimposed on a single-valued direct current. In passing through amplifiers and over the line, the direct-current component is naturally eliminated. At the TV receiver, however, the original varying direct current is needed. The clamper circuit restores the d-c component by effectively grounding the grid of the tube of the first stage of the amplifier during the horizontal synchronizing pulses of the signal. This automatically refers the entire signal to a reference base corresponding to the maximum synchronizing pulse amplitude. In other words, the output signal is held entirely above this base line instead of varying above and below a central reference. The clamping circuit employs two triodes connected as diodes, which become conducting during the synchronizing pulses that are picked off at the output of the amplifier.

The color television currently standard in the United States employs a video signal extending from a few cycles to about 4.2 megacycles. This is necessary because the color or chrominance information of the signal is modulated on a subcarrier whose frequency is 3.579545 megacycles.
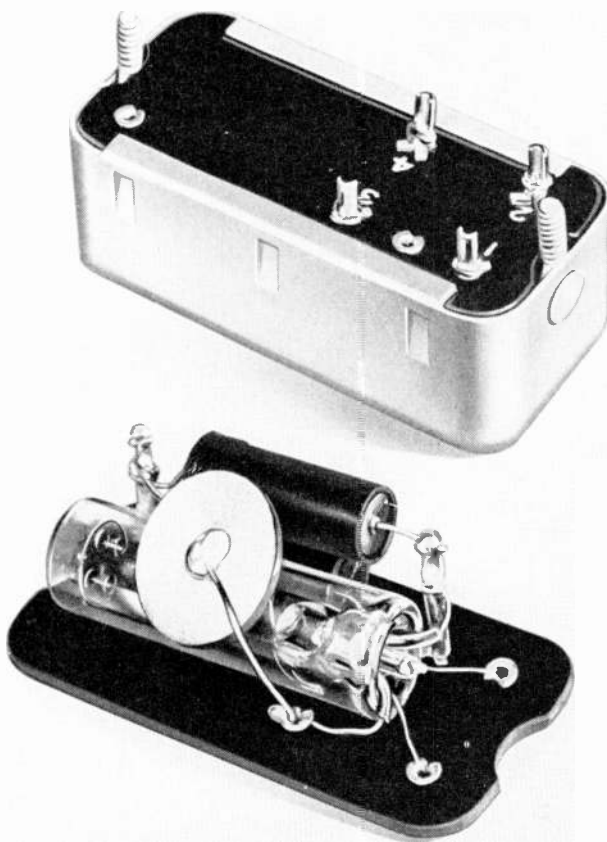
[304]

Such a frequency band is too broad for transmission over ordinary L1 facilities. Color transmission in this case requires the use of an additional modulation step which effectively shifts the color subcarrier down to a value of 2.612 mc. This results in some degradation of the "luminance" or black-and-white part of the signal which, however, has little visible effect in a color picture. When a purely black-and-white signal is sent over the same line, an automatic switch changes the filters so as to permit the signal to occupy the full frequency band.

Transmission of an unmodified color signal is well within the capacity of L3 systems. Here, the 0-4.2 mc color signal is modulated with a 4139 kc carrier so that it appears on the L3 line as an upper sideband extending between the carrier frequency and about 8340 kc. A vestigial lower sideband extending downward to approximately 3640 kc is also transmitted. This still leaves room for one master group of 600 telephone circuits in the 564 to 3084 kc band. At the receiving end the video signal is restored to its original 0-4200 kc band by modulation with the same carrier frequency of 4139 kc. The demodulating carrier is generated locally but is controlled by synchronizing pulses transmitted along with the video signal.

30.3 Type-N Carrier System

The Type-N carrier system is designed to provide 12 two-way telephone channels on two non-loaded pairs of toll or exchange cables over distances from as little as 15 or 20 miles up to a nominal maximum of 200 miles. It operates on a 4-wire basis, and uses different frequency ranges on the line for transmission in the two directions, thus making possible the use of two pairs in the same cable. In order that carrier transmission over such short distances might be economically feasible, a number of interesting innovations have been introduced to keep the cost of terminal equipment at a minimum.

The system employs throughout miniaturized equipment components mounted compactly in lightweight plug-in unit assemblies providing for maximum flexibility and ease of maintenance. Devices known as "compandors" are employed in the channel units to compress the range of speech volume on the line and thus obtain a better signal to noise and crosstalk ratio. This eliminates the need for special crosstalk balancing and noise treatment of the cable pairs. The system uses

double side-band transmission about carriers spaced at 8 kc intervals to give channel bandwidths of about 250 to 3100 cycles. Double side-band transmission permits the use of relatively simple receiving channel band filters, and eliminates the need for transmitting channel band filters altogether. The channel carriers are transmitted on the line and serve to control the regulation of repeaters, thus obviating the need of a pilot for regulating purposes. A technique known as "frequency frogging" is employed whereby the frequency groups in each direction of transmission are transposed and reversed at each repeater so that all repeater outputs are always in one frequency band, and all repeater inputs in the other. This minimizes the possibility of "interaction crosstalk" around the repeaters through paralleling voice-frequency cable pairs. The reversal of the channel groups at each repeater provides automatic self-equalization. The N-system also includes an arrangement for the transmission of dial pulses and supervisory signals, in which the
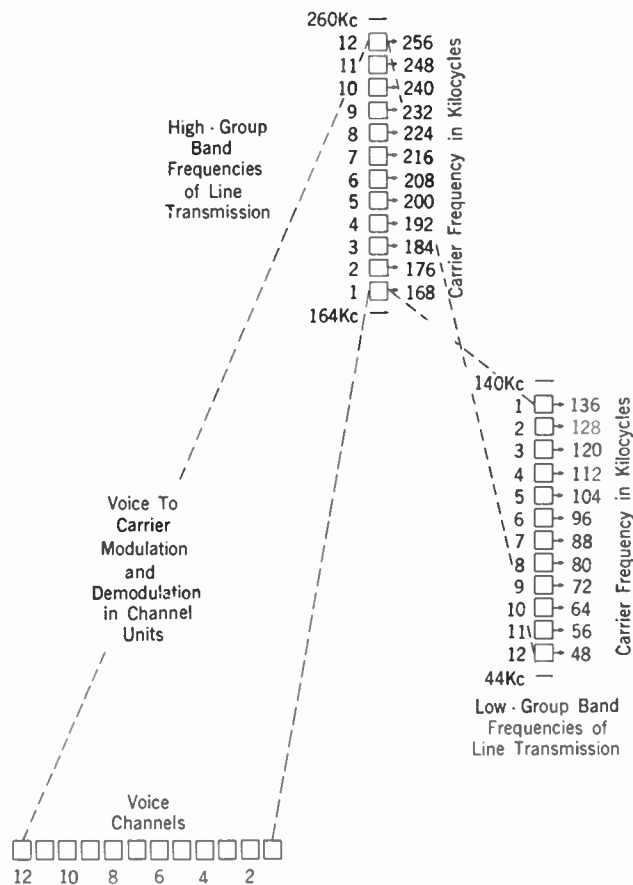
FIG. 30-11 TYPE-N CARRIER FREQUENCY TRANSLATIONS
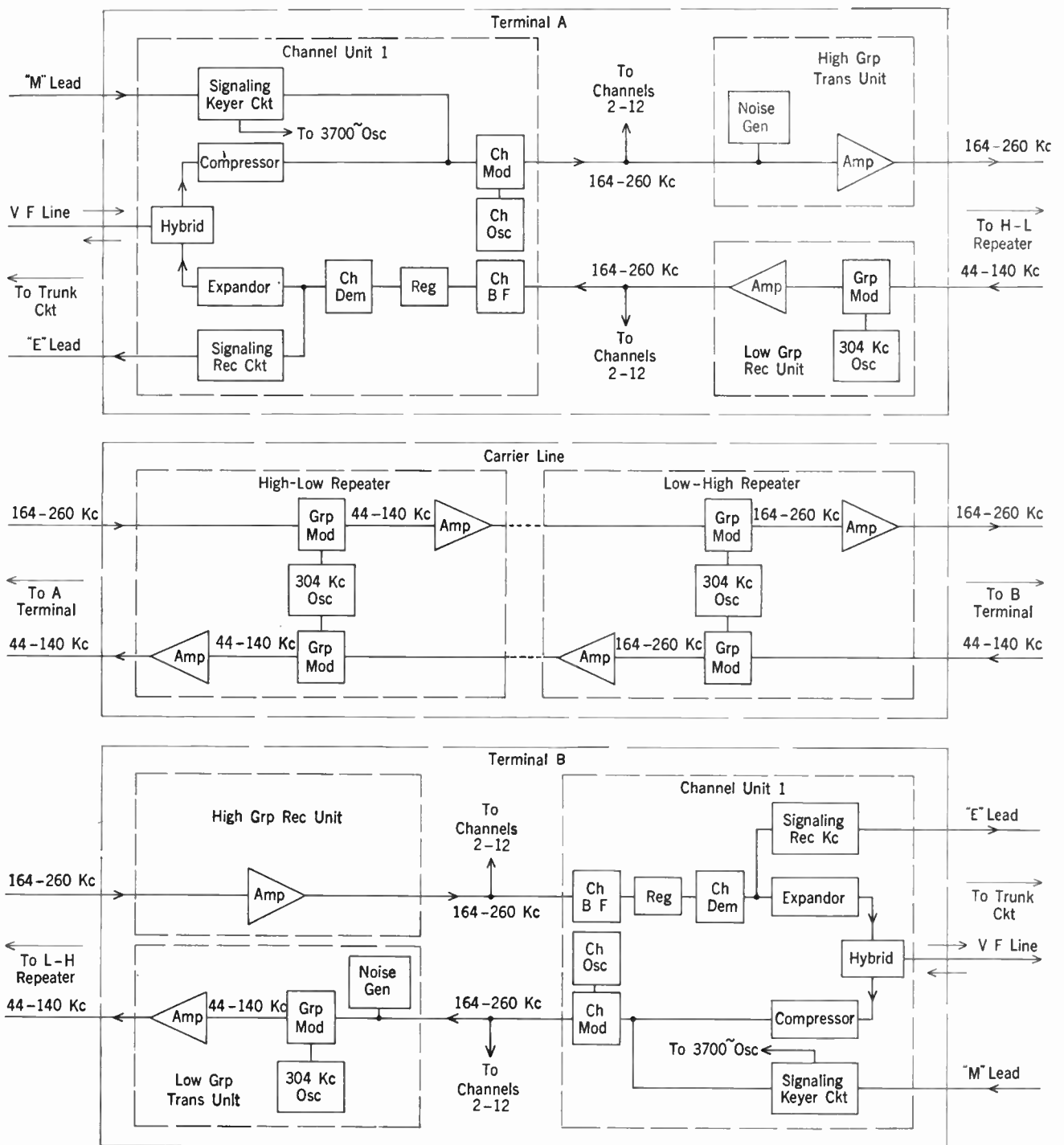
FIG. 30-12 BLOCK SCHEMATIC OF TYPE-N CARRIER SYSTEM

signaling equipment is built-in as a component part of the terminal channel unit.

The channel frequency allocations used are indicated in Figure 30-11. The channel modulators and demodulators at the terminals use only the high group of carriers between 168 and 256 kc. But either this high group or the low group may be placed on the line, depending on the number of repeater sections and other factors. It is accordingly necessary to arrange each group ter-

minal so that it can transmit or receive either high-group or low-group frequency bands. The four necessary group units are designated HGT, LGT, HGR and LGR. A high group transmitting unit always works in conjunction with a low group receiving unit and vice versa.

The general layout of the overall system is shown in block schematic in Figure 30-12. Here at terminal A, the high group is transmitted, and the HGT unit consists only of an amplifier and, when required, a controlled source of noise which may be introduced to over-ride crosstalk that might otherwise be heard on unusually quiet circuits. The incoming transmission at this terminal will be in the low group band. The LGR unit must therefore include a group modulator employing a 304 kc carrier to translate the incoming transmission to the 164-260 kc band before it is amplified and transmitted to the 12 channel units. At terminal B, on the other hand, the incoming reception in this case is in the high-frequency band and the receiving unit (HGR) consists only of an amplifier. The transmitting unit (LGT) here, however, includes a group modulator to translate the high-frequency band to the 44-140 kc band.

Repeaters along the carrier line alternate between "high-low" and "low-high". Each repeater includes two group modulators for making the necessary translations from high group to low group and vice versa. Each such translation also reverses the location of the channels in the group bands, as indicated in Figure 30-13. There reversals automatically balance out the slope char-

acteristic of two adjacent line sections to give a very nearly constant net loss across the 12-channel group. When there are an even number of line repeater sections, equalization is therefore practically complete. When the total number of sections is odd, equalization for the odd section may be taken care of by an appropriate equalizer at one terminal. The frequency reversals also reduce the maximum repeater gain to that required at the mean frequency instead of the highest frequency, and permit the use of two-stage rather than three-stage amplifiers.

With frequency frogging, the repeater spacings required for opposite directions of transmission are identical, which permits wider spacing than would be possible if transmission were in the high frequency group over the entire line. Normal repeater spacing for N-systems on 19-gage conductors is about 8 miles but must of course be less than that when smaller conductors are used. Power for the operation of repeaters can be supplied from a repeater office to an adjacent repeater on either side by a simplex connection over the line. Such repeaters may be arranged for pole mounting in cabinets. All amplifiers used in the system employ two stages with stabilized feedback. A thermistor is inserted in the feedback of the repeater amplifiers and the group receiving amplifiers to provide automatic regulation against line temperature variations. The resistance of these thermistors is controlled by the total value of the 12 channel carriers, whose power is at a much higher level than that of the voice currents. All thermistors are guarded against changes in the local ambient temperature by associated disc-type thermistors. The group modulators are of the double-balanced type employing copper-oxide disc varistors.

In addition to the compandor and signaling circuits, which are discussed later, the channel units include channel modulators and their associated oscillators on the transmitting side, and a channel band filter, regulator, and demodulator on the
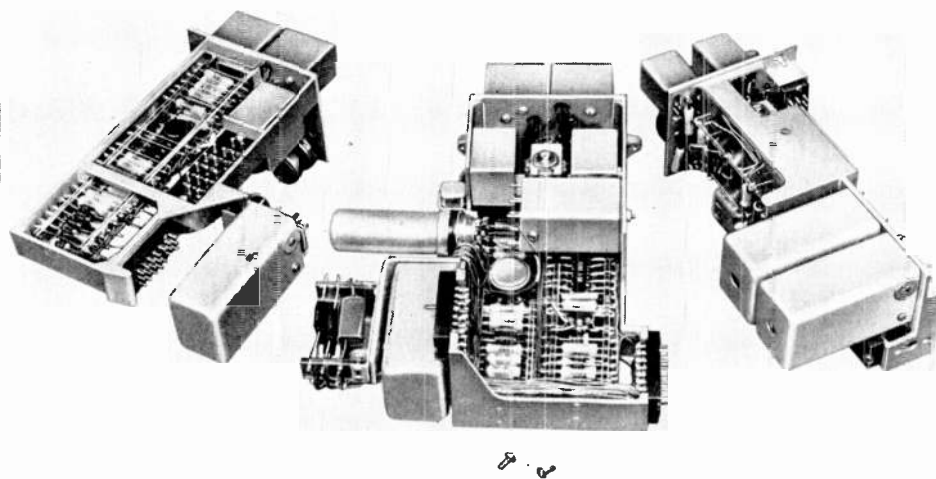
TYPE-N CARRIER UNIT SUBASSEMBLIES. COMPRESSOR AT LEFT; EXPANDOR AND SIGNALING AT CENTER; CARRIER AT RIGHT

receiving side. The channel modulator is a bridge type arrangement of point-contact germanium varistors with its carrier supplied from a crystal controlled electron-coupled oscillator. In the receiving side of the circuit, the regulator is a two-stage amplifier arranged to provide delayed automatic volume control by the application to the grid of the input tube of rectified carrier picked off at the output of the demodulator. The demodulator is a lattice arrangement of germanium varistors, which operates as a linear detector.

As Figure 30-12 indicates, the **compandor** consists of two separate devices—a **compressor** and an **expandor**. Devices of this type have sometimes been used for noise reduction in long distance radio circuits and in Type-J carrier circuits, but are incorporated as an integral part of a system for the first time in Type-N carrier. The compressor provides a variable gain, the value of which depends upon the strength of the incoming voice signal, and the expandor provides a variable loss which depends upon the same factor. The overall action of the two devices is illustrated in Figure 30-14 in terms of the signal levels. From this, it may be noted that a signal of $+5$ dbm is applied to the line without modification by the compressor. A signal input of -50 dbm, however, receives 27.5 db gain in the compressor and
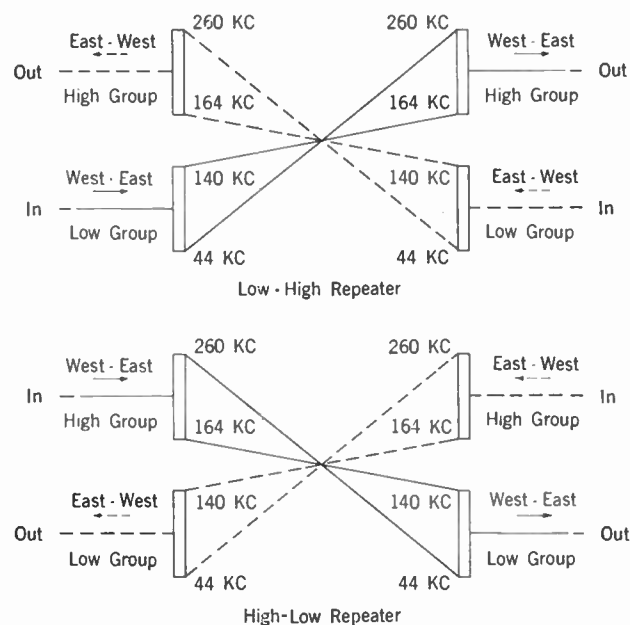
Fig. 30-13 Frequency Translations at Repeaters

is applied to the line at -22.5 dbm. The expandor in turn introduces a loss of 27.5 db to return the signal to its original level. At intermediate values of input, the compression and expansion is proportionately smaller. For each signal input
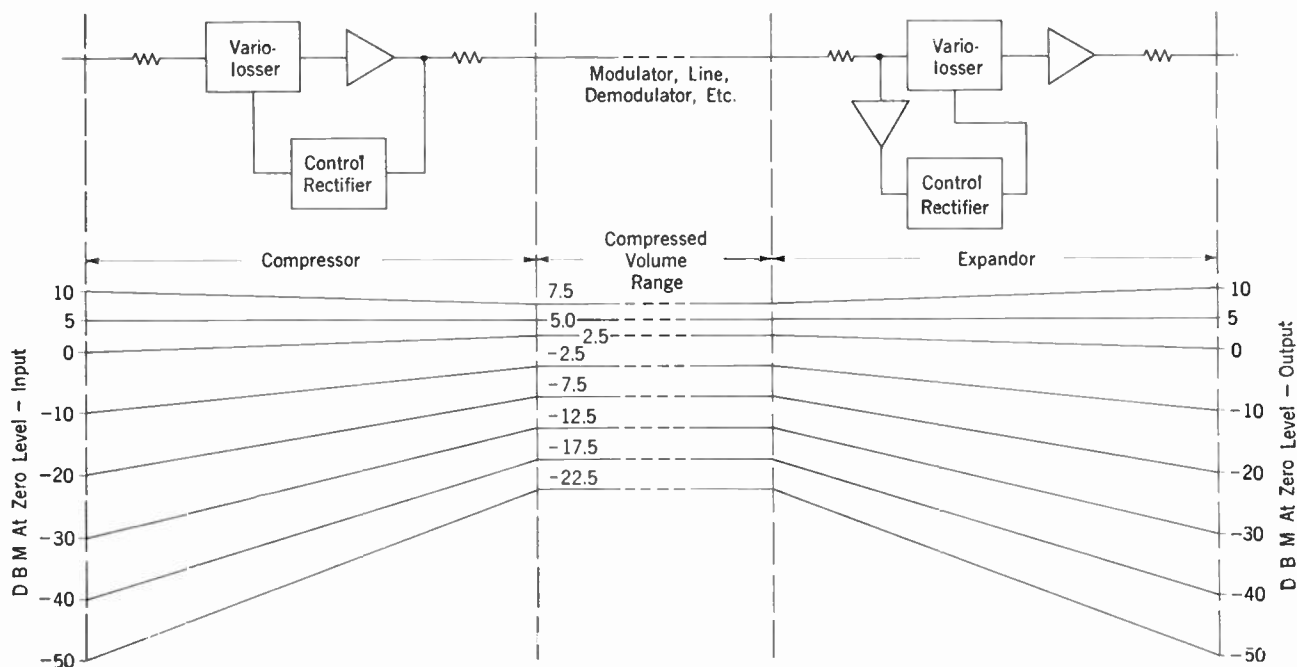
Fig. 30-14 Compandor Action on Signal Levels

[308]

weaker by 2 db than +5 dbm, the compressor introduces 1 db more of gain and the expandor 1 db more of loss. In other words, the compression ratio is 2 to 1.

Schematics of the compressor and expandor circuits are shown in Figure 30-15. The overall gain of the compressor circuit is determined by the variolosser which precedes the two-stage amplifier. This is effectively a resistance pad having two shunt arms whose resistance values vary in accordance with the value of direct current flowing through the germanium varistors of which they are composed. The direct current is obtained from the control circuit, which rectifies a portion of the amplifier output voltage. Thus, if the amplifier output rises, the direct current in the variolosser varistors increases, which increases its attenuation so that the input to the amplifier decreases. The design is such that a 2 db change in the variolosser input results in only 1 db change in the variolosser output. The control circuit contains a filter, whose time constants are such that the direct current varies in accordance with the syllabic envelope of the speech in the amplifier output. The expandor circuit operates on the same general principles as the compressor. Its control circuit, however, is operated by the input speech signal rather than by the output signal.

The built-in signaling circuits of the N-system are designed primarily to handle dial pulse and supervisory signals. Ringdown signals at 1000 cycles and multi-frequency key pulsing signals are passed directly over the voice channel. Dial pulses and supervisory signals are transmitted by turning on and off a 3700 cycle signaling tone which is injected at the transmitting terminal following the compressor, and removed at the receiving end before reaching the expandor. The 3700 cycle tones after modulation appear on the line as side frequencies 3700 cycles above and below the channel carrier frequencies. They are kept separated from the voice channels, which extend only up to 3100 cycles, by appropriate low-pass and high-pass filters (not shown in Figure 30-12).

Figure 30-16 is a schematic of the signaling keyer circuit, which is included in each transmitting branch of each channel unit. It includes two germanium varistors which act to transmit, or to block, the 3700-cycle oscillator output. Thus for supervisory signaling, when ground is connected to the "M" lead from the trunk, positive potential causes current to flow in the varistors in such a

direction as to make their resistance low and permit the transmission of signal tone to the channel modulator. In the off-hook condition, on the other hand, the negative potential from the −48 volt battery is controlling. This causes the varistors to have high resistance, blocking the transmission of the signal tone.

On the receiving side, the 3700-cycle tone is picked off after the channel demodulator and applied to the signaling receiver circuit shown

Type-N Channel Units—Front View

Voice Amplifier

Variolosser

Inter-stage Net-work

− 130 V

Filter

Rectifier

Pad

Control Circuit

Compressor Circuit

Output Amplifier

Input Transformer

Output Transformer

Variolosser

+130 V

+130 V

Input Transformer

Rectifier

Control Varistors

− 40 V

+130 V

Output Transformer

Control Amplifier

Expandor Circuit
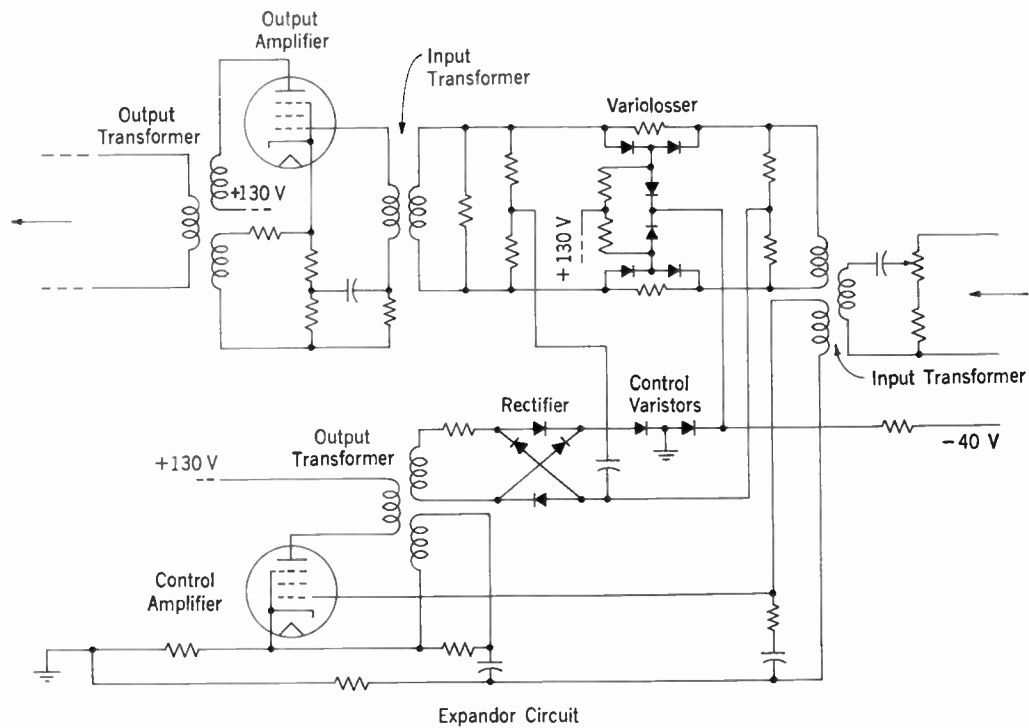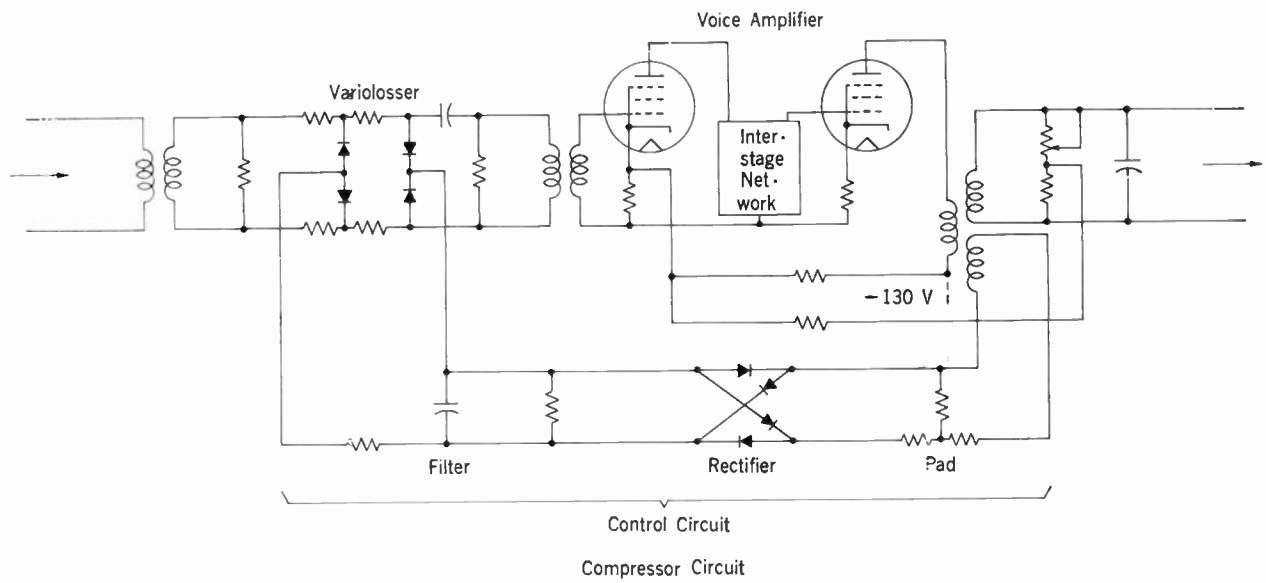
FIG. 30-15 COMPANDOR SCHEMATICS

schematically in Figure 30-17. After selection by the band-pass filter, the 3700-cycle signal is amplified and applied to a limiter-multivibrator that transforms the sine wave signal, which may vary somewhat in amplitude, into a square wave signal of constant amplitude. This reshaping of the signal wave prevents the time-delay circuit, which follows, from introducing pulse width distortion that might cause erroneous registration of dial pulses. The output of the limiter-multivibrator is fed through an impedance-matching cathode-follower to a voltage-doubling rectifier whose output is connected through a delay network to the grid of a "d-c amplifier". The time constant of

[310]

the delay network is such as to make the overall response time of the circuit 15 to 20 milliseconds, which prevents short noise bursts or other transients from causing false operation. The plate circuit of the d-c amplifier is connected to the operating winding of a mercury-contact polarized relay, which in operating disconnects ground from the "E" lead to the switchboard trunk. When no 3700-cycle tone is being received, the d-c amplifier tube is biased by −40 volts, no plate current flows, and the pulse-relay is held in its non-operated position by the current in its biasing winding. When 3700-cycle tone is present, the rectifier output voltage overcomes the tube bias permitting the flow of a large plate current. This operates the pulse-relay and removes ground from the "E" lead.
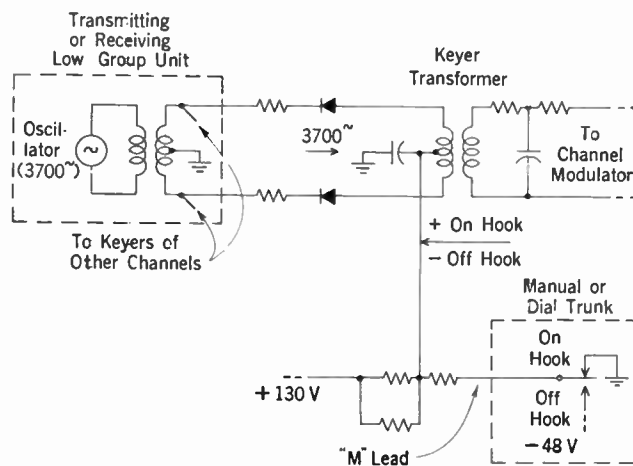
FIG. 30-16 SIGNALING KEYER CIRCUIT

30.4 Types O and ON Carrier Systems

The Type-O carrier system was designed to provide relatively short-haul carrier channels over open wire conductors. It makes use of miniaturized equipment and many of the other features of the Type-N system including compandors, frequency-frogging and built-in 3700 cycle signaling. The system is capable of providing a maximum of 16 two-way voice channels on one pair of wires, occupying a frequency range extending from 2 to 156 kilocycles. Actually, the channels are arranged in four 4-channel groups designated OA, OB, OC, and OD. Any one or more of these groups may be separately employed to obtain additional circuits as required in integer multiples of four
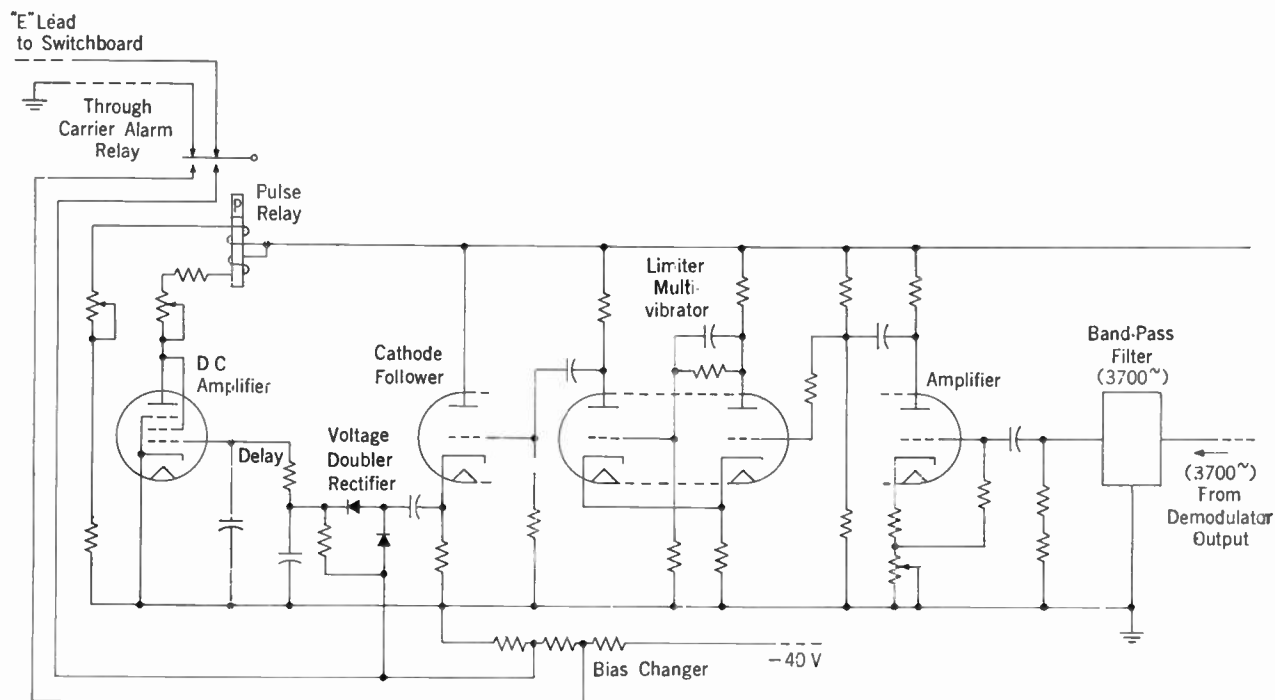
FIG. 30-17 SIGNALING RECEIVER CIRCUIT
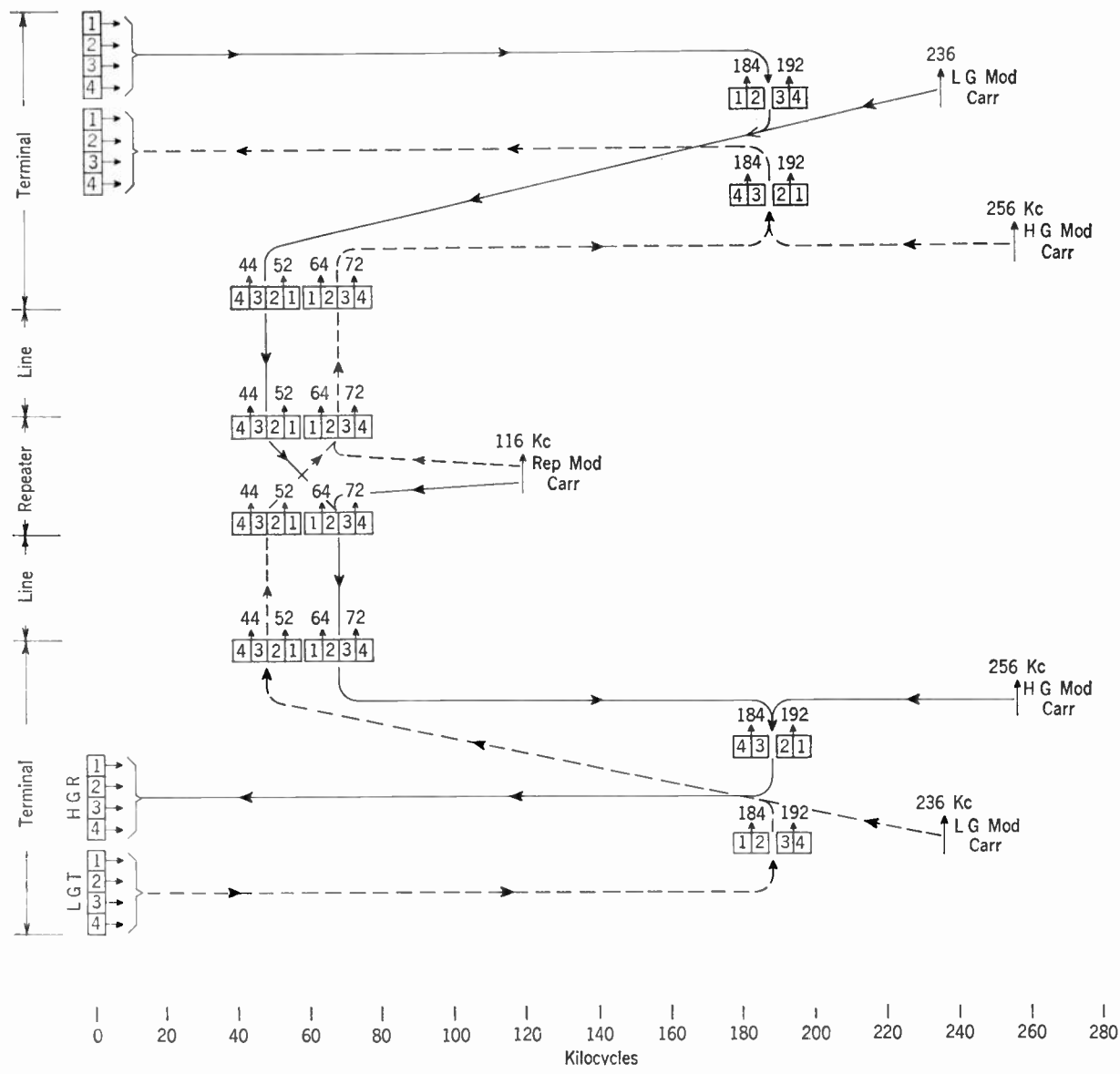
FIG. 30-18 TYPE-OB1 CARRIER FREQUENCY TRANSLATIONS

up to the maximum of 16. The line frequencies employed are as follows:

Type	Low Group	High Group
OA	2-18 kc	20-36 kc
OB	40-56	60-76
OC	80-96	100-116
OD	120-136	140-156

As in the N-system, terminals are arranged to transmit either a low or high group of frequencies, and to receive the corresponding opposites. Re-

peaters, which are spaced at intervals of 40 to 50 db, are also arranged alternately for low-high or high-low transmission. Unlike the N-system, however, single side-band transmission is used in order to conserve frequency space. Here one voice channel and a carrier are supplied to a modulator from whose output everything but one sideband is filtered out. At the same time, another channel is modulated on the same carrier frequency and only the opposite sideband is retained. The carrier frequency is thus suppressed entirely but it is separately reinserted at a predetermined amplitude for transmission over the line along with

[312]

the two sidebands. In this way, only two carriers, spaced 8 kc apart, are required for four voice channels. The same two carriers, 192 and 184 kc, serve for the initial modulation of all of the four-channel groups (i.e., OA, OB, etc.). For transmission over the line, a further modulation step is required to place each voice channel in its appropriate frequency allocation. Thus, Figure 30-18 shows the frequency translations employed in the terminal channel and group modulators and at the repeaters for OB transmission. Different carrier frequencies must of course be used in the group modulators and repeaters of the other four-channel groups to obtain the respective line frequencies given in the table above.

In all cases, the two carriers of each system are transmitted over the line, where their combined power is used for regulation of the amplifiers at repeaters and group receiving stations to correct for line attenuation variations. Regulation is accomplished by means of a type of automatic volume control in which a part of the output of the line amplifier following the modulator is picked off and amplified in a "control amplifier", rectified and fed back to the input of a regulating amplifier that precedes the modulator. A block schematic of the arrangement of a repeater is given in Figure 30-19. Other units employed in the O-system are practically identical with their counterparts in the N-system, which have already been discussed.

There are many situations in practice where it is desirable to operate over open wire and cable facilities in tandem. For this purpose, a system known as Type-ON carrier was developed. This system provides for converting Type-O channel groups in such a way that they may be transmitted over an N line. Any five 4-channel O groups transmitting in one direction are converted to five ON groups and applied to the N cable pair transmitting in that direction. Each of the five O groups may be from an OA, OB, OC, or OD system, either high group or low group. The corresponding five
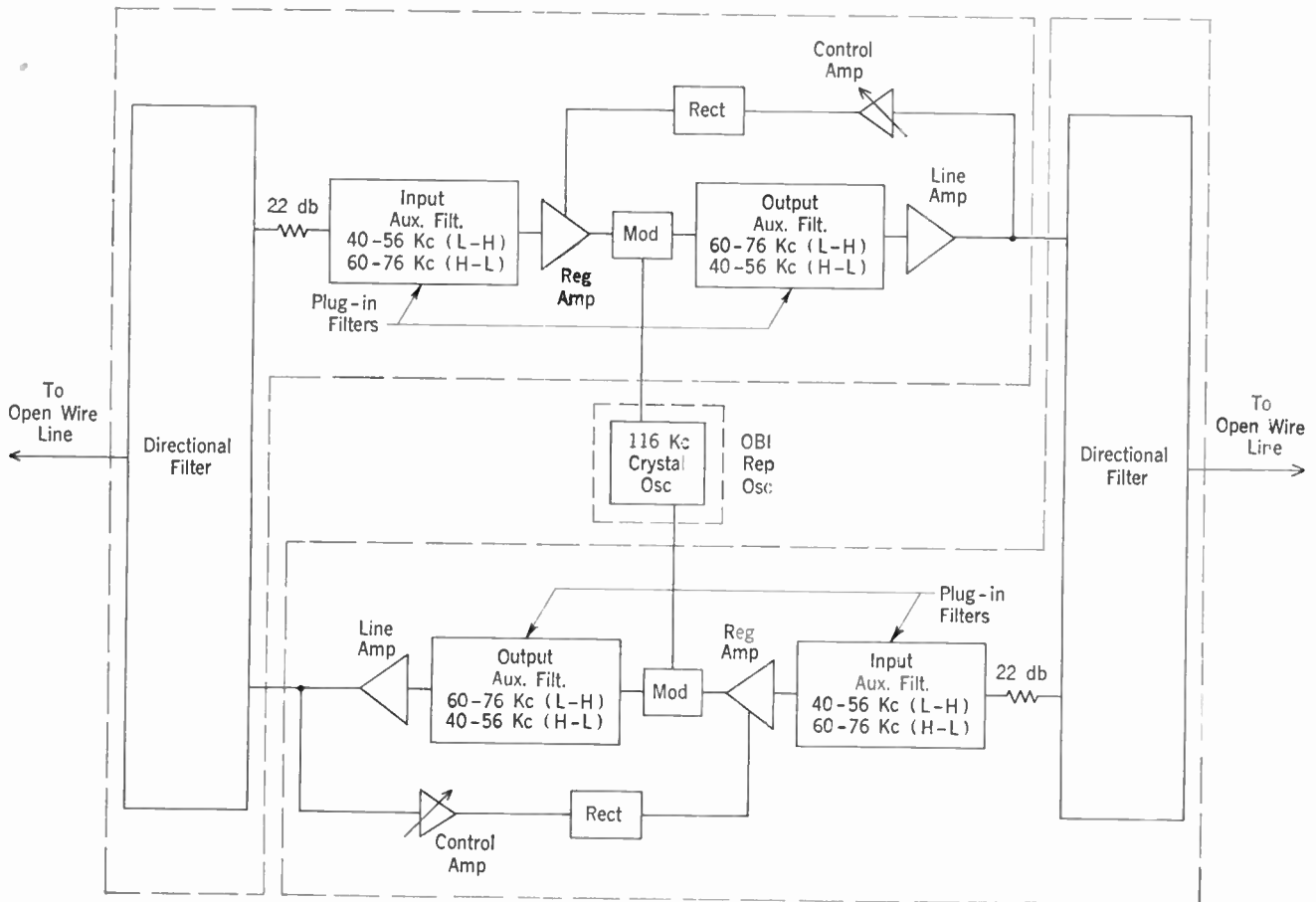
FIG. 30-19 BLOCK SCHEMATIC OF TYPE-OB1 REPEATER

groups transmitting in the opposite direction are likewise converted to ON groups and connected to an N cable pair transmitting in the reverse direction. To accomplish this, each of the ten O groups is first translated to the basic ON 4-channel group frequency of 180-196 kc, and then by a second modulation placed in its allotted frequency space on the N line. The five groups transmitting in one direction are thus stacked in the band between 36 and 132 kc while the opposite groups occupy the range between 172 and 268 kc.

Although the ON system was designed primarily to permit direct interconnection between O and N systems, it may also be applied to cable facilities alone in place of N-carrier. Because it employs single instead of double side-band transmission, it permits the derivation of 24 circuits within the same total frequency band that provides 12 circuits in the N-system.

CHAPTER 31

RADIO SYSTEMS

31.1 Types of Radio Systems

The essential elements of any radio system are (1) a transmitter for modulating a high-frequency carrier wave with the signal, (2) a transmitting antenna that will radiate a maximum amount of the energy of the modulated carrier wave, (3) a receiving antenna that will intercept a maxi-

mum amount of the radiated energy after its transmission through space, and (4) a receiver to select the carrier wave and detect or separate the signal from the carrier. Although the basic principles are the same in all cases, there are many different designs of radio systems. These differences depend upon the types of signal to be transmitted, the distances involved, and various other factors, including particularly the part of the frequency spectrum in which transmission is to be effected.

Figure 31-1 is a chart of the radio spectrum indicating at the left the commonly accepted classification of radio frequency ranges; and showing at the right the more important frequency ranges of special interest in current telephone practice. It will be noted that telephone practice makes use of some part of nearly all of the major frequency ranges. It must accordingly employ a corresponding variety of types of radio facility. It is not practicable or desirable to attempt to describe all of these in this book, and what follows will therefore be limited to a brief general discussion of principles applicable to all radio systems, with a few examples of specific facilities.
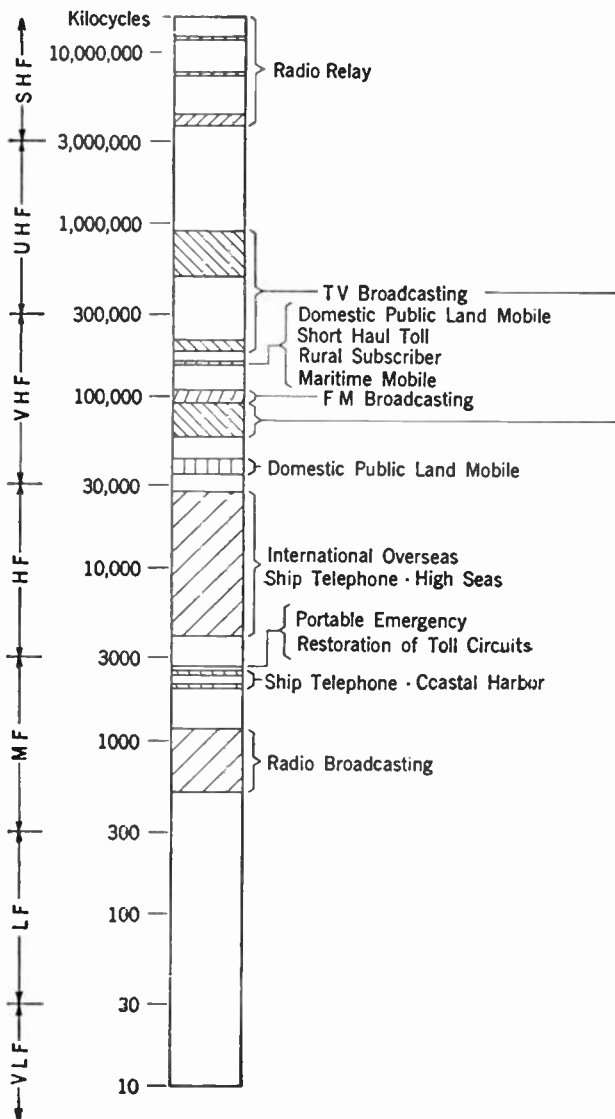
FIG. 31-1 RADIO FREQUENCY SPECTRUM
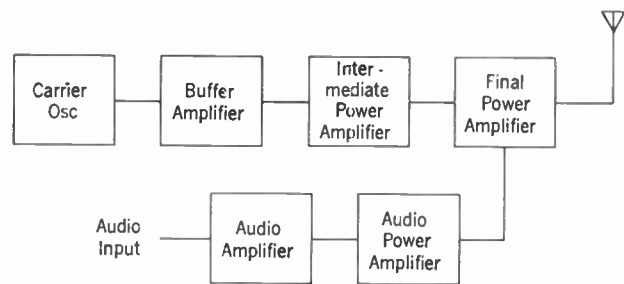
FIG. 31-2 AM RADIO TRANSMITTER—HIGH-LEVEL MODULATION

31.2 Radio Transmitters

The principal components of a typical amplitude-modulated radio transmitter, such as might be used in radio broadcasting or for relatively low-powered point-to-point transmission, are indicated in the block diagram of Figure 31-2. Here

[315]

the amplified input signal plate-modulates the carrier in the output circuit of a Class C power amplifier which represents the final stage in a chain of amplifiers that increases the power of the
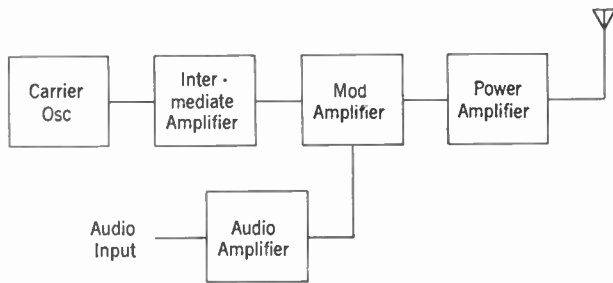
FIG. 31-3 AM RADIO TRANSMITTER–LOW-LEVEL MODULATION

carrier to an appropriate value for application to the antenna.

Another common design of transmitter for relatively low power work applies the signal to the plate circuit of an intermediate modulating amplifier, as indicated in Figure 31-3. As compared with the arrangement of Figure 31-2, this has the advantage of requiring less amplification of the audio signal to secure full modulation. On the other hand, it necessitates the use of a linear amplifier in the final power stage instead of the more efficient Class C amplifier that may be used when the modulating signal is applied to the final power stage as in Figure 31-2.

For very long distance point-to-point radio telephone circuits, such as those used in transoceanic service, transmitter design is naturally somewhat more elaborate. Figure 31-4 is a block schematic showing the major components of a transmitter (coded LD-T2) used in the Bell System for multi-channel single-sideband transmission in the 4 to 28 megacycle range. As may be seen, this transmitter employs a triple modulation system followed by a six-stage linear amplifier. It is arranged to accept two independent voice-frequency bands, each extending from 100 to 6000 cycles. These bands are broad enough to permit superimposing two separate voice channels on each by appropriate carrier methods. The two band groups are transmitted as the upper and lower side-bands of a single carrier frequency, thus providing a total of four telephone circuits on one radio channel. The carrier itself is also transmitted at reduced volume for automatic frequency control and automatic volume control at the receiver.

As Figure 31-4 indicates, the two voice-frequency bands are applied separately to two low-frequency modulators, together with a 100 kc carrier. Band-pass filters following the LF modulators select two side-bands, the upper of which extends from 100.1 to 106 kc, and the lower from 99.9 to 94 kc. The carrier is at this point eliminated. The two side-bands are combined in a hybrid and applied to a two-stage amplifier at the output of which the 100 kc carrier is re-supplied
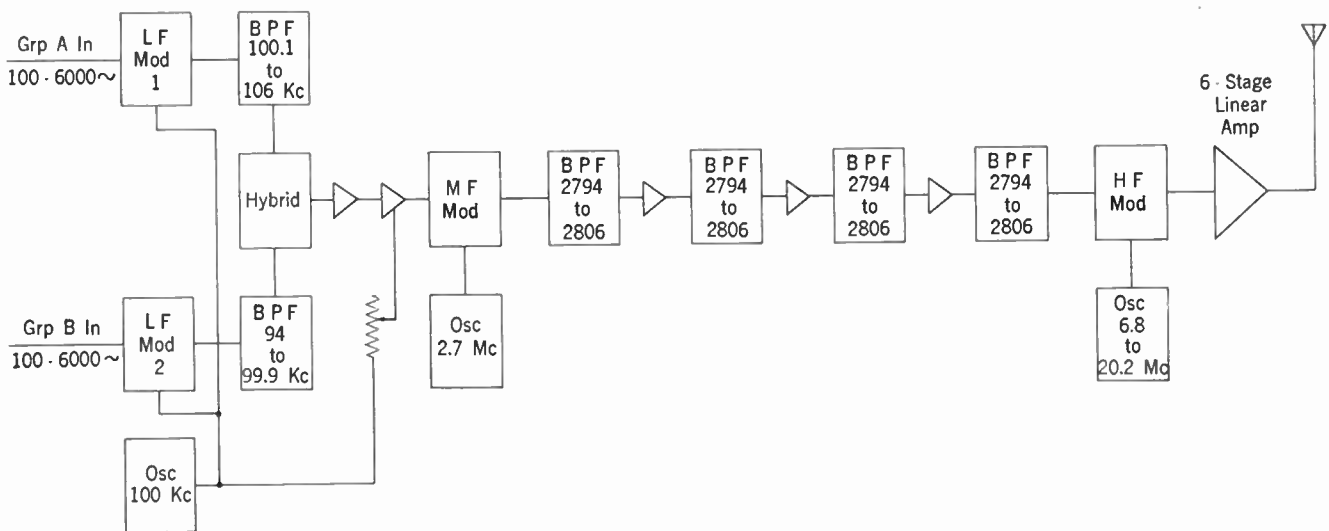
FIG. 31-4 LD-T2 RADIO TRANSMITTER

[316]
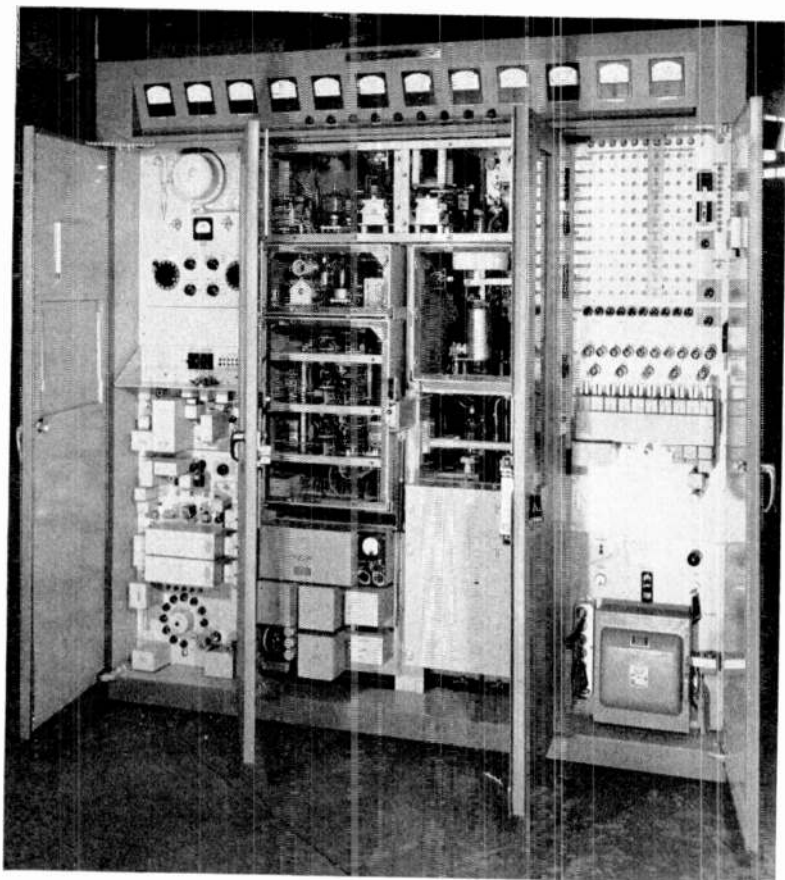
LD-T2 Radio Transmitter (Front View)

at reduced volume. A second, medium-frequency modulator stage translates the total signal to an intermediate frequency range extending from 2794 to 2806 kc, which is the upper side-band of a 2700 kc conversion frequency. The original 100 kc carrier is now at 2800 kc. Precision of signal range is secured by a series of four 2794 to 2806 kc band-pass filters working in tandem with three medium-frequency amplifiers.

The final frequency conversion takes place in the high-frequency modulator where the signal centered about 2800 kc modulates a conversion frequency, which may have any one of a number of values ranging between 6.8 and 20.2 megacycles. These frequencies are supplied by an oscillator circuit whose output frequency is determined by the controlling crystal that is connected into the circuit. The transmitter is equipped with a set of crystals which provide ten different frequencies. After final modulation, the signal wave is applied to a six-stage linear amplifier which is capable of delivering a peak power of 4 kw. The first four stages of the amplifier employ pentodes, the fifth stage tetrodes, and the final stage a 2.5 kw air-cooled triode operating with grounded grid.

As was pointed out earlier, there are a number of different types of circuits in use for frequency modulating a radio carrier wave. Figure 31-5 indicates in block schematic the arrangement of an FM transmitter in which modulation is effected by means of a reactance tube circuit. Because FM radio (voice) transmission is generally in the very high-frequency range, it is necessary to use frequency multiplying circuits to bring the basic frequency generated by the master oscillator up to the desired value. The oscillator usually operates in the neighborhood of 5 mc, and its output frequency must be multiplied by factors in the order of 10 to 20 to reach the frequencies prescribed for FM transmission. Two or three frequency-multiplication stages are usually employed
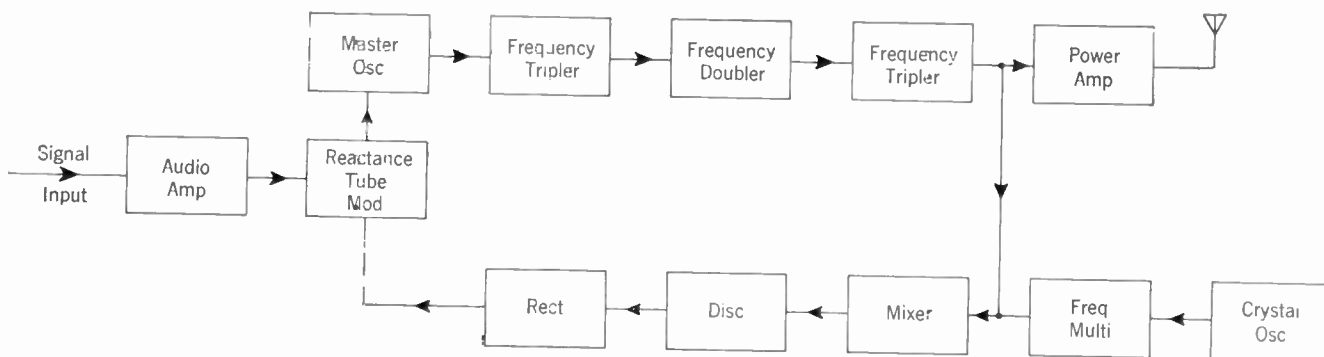
Fig. 31-5 FM Transmitter Using Reactance Tube

[317]
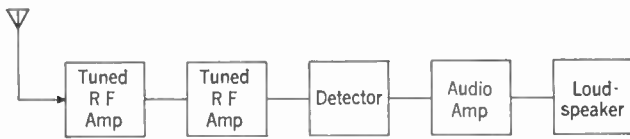
FIG. 31-6 TRF RADIO RECEIVER

for this purpose.

In order to maintain the carrier frequency at a fixed value, the reactance tube type of transmitter requires the use of an automatic frequency control arrangement as shown in the lower part of Figure 31-5. This includes a crystal-controlled oscillator of highly stable frequency. A portion of the modulated carrier is picked off and compared in a mixer (modulator) circuit with the output of the crystal oscillator. The difference between the two frequencies, if any, is fed to a discriminator, the output of which, after rectification, is applied to the grid of the reactance tube. The polarity of this rectified output will be such as to hold the mean frequency of the master oscillator effectively constant under the control of the crystal oscillator.

31.3 Radio Receivers

The radio signals that are picked up by the antenna of a radio receiver are usually very weak so that the receiver circuit must ordinarily include one or more amplifiers. Figure 31-6 is a block schematic of a tuned radio-frequency (TRF) receiver for amplitude-modulated signals. With this arrangement, the r-f signal taken from the antenna passes through from one to three or more stages of amplification before reaching the detector. Associated with the input of each amplifier is a tank circuit which can be tuned by a variable capacitor to the frequency of the desired incoming signal. After detection, one or more stages of
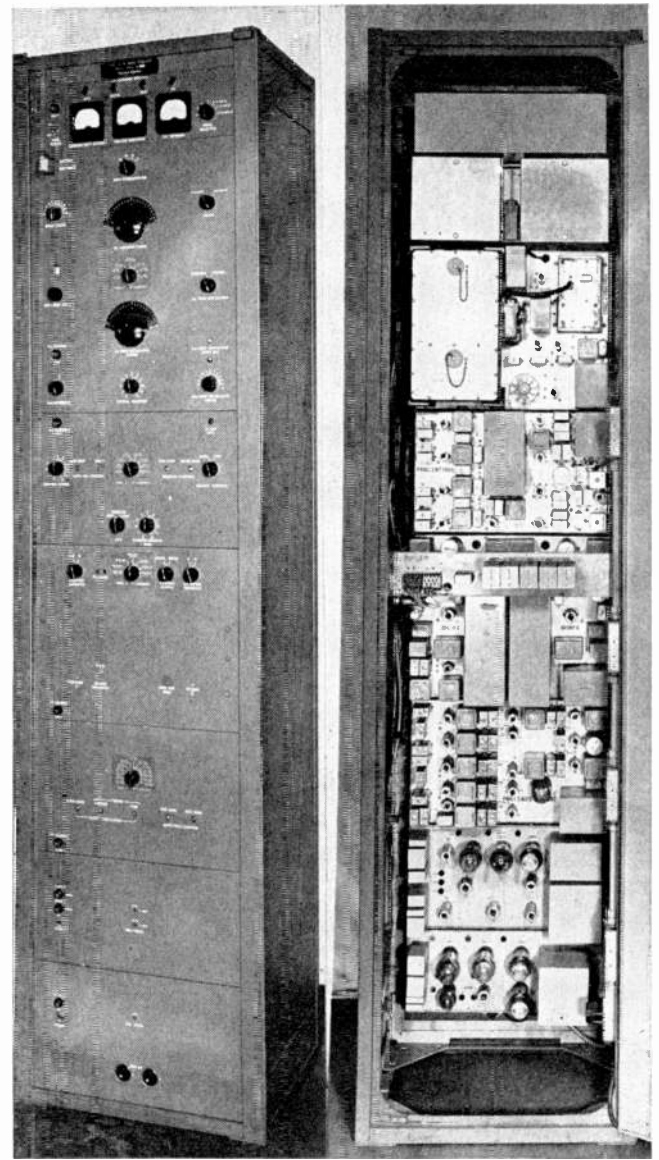
(FRONT)   (REAR)

LD-R1 RADIO RECEIVER

audio amplification are employed to bring the detected signal up to the strength required to operate the loud-speaker or other sound-reproducing device. TRF receivers, while once in general use, are used at the present time only on a limited basis, chiefly because of the difficulty of securing sufficiently sharp selectivity with respect to adjacent channels.
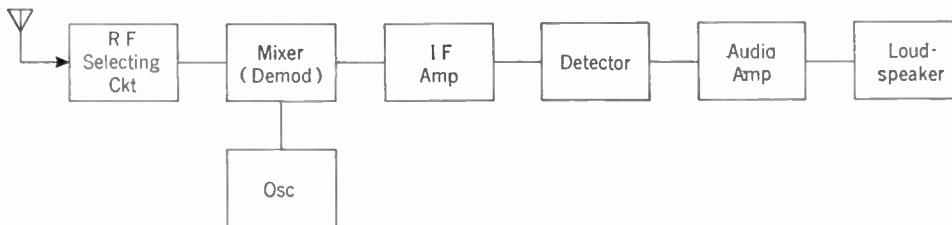
FIG. 31-7 SUPERHETERODYNE RADIO RECEIVER

The vast majority of modern receivers are of the "superheterodyne" type illustrated in block schematic in Figure 31-7. Here, before detection, the r-f signal is converted by a modulation process to a fixed intermediate frequency value in which most of the required amplification takes place. The incoming signal is selected by a variable tuned circuit, which in some cases may include one amplification stage. A local oscillator, which supplies the mixer or demodulator (also sometimes called a converter or first detector), is tuned simultaneously with the signal selecting tuner so that the frequency of the mixer output, which is the difference between the frequencies of the oscillator and the incoming carrier, is always the same. The intermediate frequency amplifier circuits, accordingly, require no adjustment and may employ coupled circuits double-tuned to a single constant frequency. Radio broadcast receivers are usually designed for an intermediate frequency value of about 450 kc. Figure 31-7 indicates a local oscillator separate from the mixer circuit, but in most broadcast receivers a single multi-grid tube known as a "pentagrid converter" performs both the oscillator and mixer functions.

The superheterodyne receiver, although its advantages are sufficient to warrant its general use, has some inherent capacity to produce spurious responses under certain conditions. One of the major sources of such undesired responses is the possible presence at the antenna of a signal whose
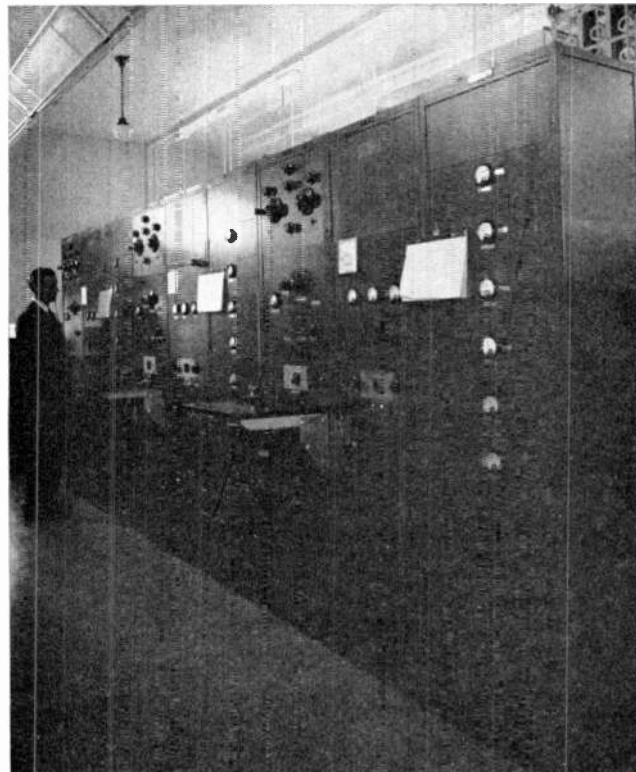
SINGLE SIDEBAND OVERSEAS RADIO RECEIVERS

value is the "image frequency" of the tuned-in signal. The frequency of the image signal is greater than the frequency to which the receiver is tuned by twice the value of the intermediate frequency
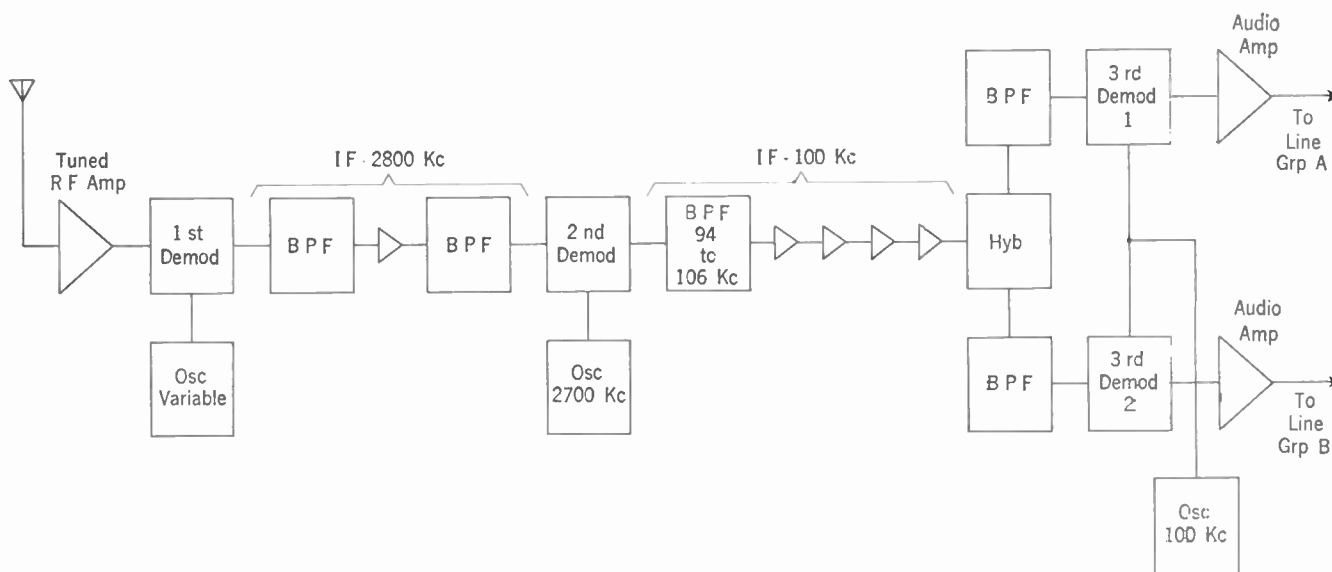
FIG. 31-8 LD-R1 SINGLE SIDEBAND RADIO RECEIVER
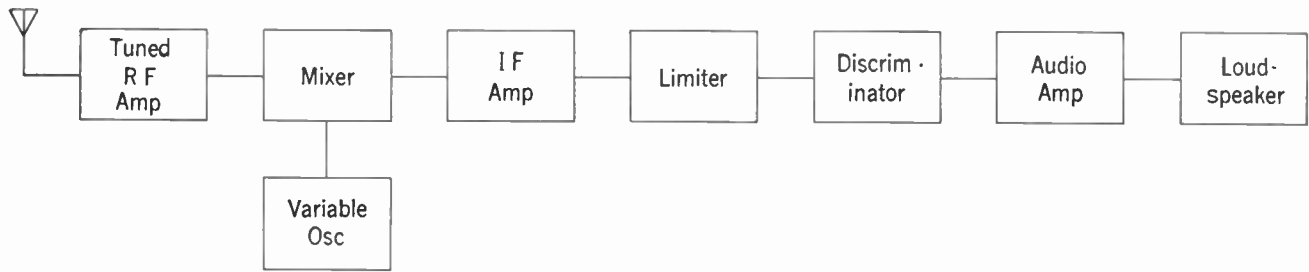
FIG. 31-9 FREQUENCY-MODULATION RECEIVER

if, as is normally the case, the local oscillator is operating at a higher frequency than that of the desired signal. Such an image signal will mix with the oscillator frequency to produce a difference frequency that is exactly equal to the intermediate frequency. Thus, both signals would be amplified in the IF section and appear simultaneously in the receiver output.

Receiver response to image signals can only be avoided by blocking the image in the selecting circuits that precede the mixer. This is facilitated by the use of an intermediate frequency which will cause the image of the desired signal frequency to lie at a considerable distance in the frequency band from the desired frequency. For example, if the intermediate frequency is 450 kc,

and the receiver is tuned to a signal carrier at 800 kc, the oscillator frequency must be 1250 kc. With this oscillator frequency, the image signal that would produce a 450 kc intermediate frequency would be 1250 kc plus 450 kc or 1700 kc. This is sufficiently removed from the 800 kc signal so that little, if any, is likely to pass through the tuned selecting circuit to reach the mixer.

For reception of long distance point-to-point signals, as in transoceanic service, a more elaborate receiver employing two intermediate frequencies is often used. This arrangement is sometimes known as a triple-detection receiver. The first intermediate frequency has a relatively high value to permit maximum image suppression, and the second intermediate frequency is comparatively low to provide high adjacent channel selectivity.

The Bell System single-sideband receiver coded LD-R1, which is shown in block schematic in Figure 31-8, is of this general type. This receiver operates in the frequency range from 4 to 28 mc. After the first demodulation, the signal is centered around an IF of 2800 kc, a frequency high enough to give good suppression of the image 5600 kc away from the desired signal. The band-pass filters on each side of the amplifier have high rejection characteristics for frequencies 200 kc below

OVERSEAS HF RADIO TRANSMITTERS

[320]

2800 kc in order to avoid the possible transmission of a second image frequency into the 100 kc section of the circuit.

The second demodulator, which is supplied from a 2700 kc oscillator, translates the signal into the 100 kc range, a band-pass filter selecting frequencies between 94 and 106 kc. The main four-stage amplifier delivers the signal through a hybrid to two band-pass filters which select the upper and lower side-bands of the 100 kc carrier. Each side-band group is then converted to voice frequencies by a balanced copper-oxide demodulator supplied with 100 kc from a local oscillator. Each group is sufficiently wide to provide for two telephone channels, as in the case of the comparable transmitter previously discussed.

The pre-detection stages of receivers for FM signals are generally the same as those of AM receivers. As shown in Figure 31-9, however, detection in this case is effected by means of limiting and discriminating circuits. For best results, IF amplification should be great enough to raise all peaks of the signal above the cut-off point of the limiter. This will automatically eliminate any amplitude variations that may be present so that the signal at the output of the limiter will have a uniform fixed amplitude and will vary only in frequency. By eliminating amplitude variations, noise and unwanted energy which are the chief causes of such variations are minimized. The discriminator converts the constant amplitude frequency-variations into an audio-frequency signal, as described in Article 26.3, which is then amplified in the usual manner.

31.4 Antennas

The effectiveness of a transmitting antenna is measured by its ability to convert a maximum amount of the power developed by a radio transmitter into radiant energy in the form of electromagnetic waves, which will be transmitted in such a direction as to produce maximum field strength at the receiver. In point-to-point transmission, and to some extent in broadcast transmission, the factor of major importance is usually the degree of directivity or antenna gain (Article 17.4) that can be secured. Transmitting and receiving antennas are sometimes alike but there are many situations where economic considerations and other factors require quite different designs for the two conditions.

There are many different antenna designs and arrangements in practical use in radio work. The various designs may be grouped into a limited number of major types, however, in which the controlling design factors are the frequency range in which the antenna is to operate and the degree of directivity desired. In general, antenna effectiveness can be more economically increased as the frequency of the radio wave is increased.
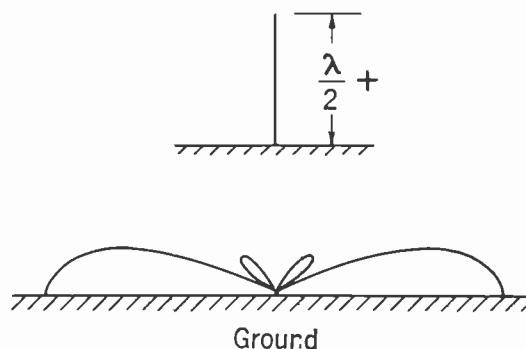
FIG. 31-10 RADIATION PATTERN OF HALF-WAVE VERTICAL ANTENNA

Transmitting antennas for medium radio frequencies quite commonly employ a simple vertical radiator which may consist of a single wire or a slender steel tower. Such radiators are tuned to resonance for the carrier frequency to be transmitted by making their total height equal to an appropriate fraction of the carrier wavelength; or, if the wavelength is too long to make this practicable, by adding lumped reactance in series with the radiator at its top (capacitive) or base (inductive). In either case, the antenna when energized at its resonant frequency, behaves like a resonant transmission line with a standing wave of current extending along the conductor.

The typical radio broadcast antenna is a steel tower slightly more than one-half wavelength in height and effectively grounded at its base where the energy is applied. The vertical radiation pattern of such an antenna is shown in Figure 31-10, where the lobes indicate the relative field strengths radiated in various directions. The Figure of course represents the pattern of a single vertical plane through the antenna and the total pattern would be represented by rotating this plane through 360° about the antenna. This antenna,

[321]

accordingly, radiates with equal strength in all horizontal directions but concentrates the field along the horizontal, with relatively low radiation upward at high angles. Both of these features are generally desirable for ordinary broadcasting purposes. Where, as may be the case even in broadcasting work, some directivity in azimuth (horizontal) is desired, one or more additional radiators may be employed. The geometrical arrangement and spacing of the additional radiators and the phase relationships of the currents with which they are driven will determine the direction of maximum radiation.

In the high-frequency range (3 to 30 mc) extensive use is made of "long-wire" antennas for point-to-point transmission. A single horizontal conductor mounted a half-wavelength or more above the ground is directional along the axis of the wire when the wire is several wavelengths long. Thus, Figure 31-11 shows the radiation pattern of an isolated resonant wire of length 5λ. While the pattern includes numerous minor "spurious" lobes, the axes of the major lobes are at relatively small angles to the antenna, and
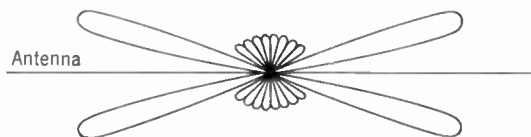
FIG. 31-11 RADIATION PATTERN OF LONG-WIRE ANTENNA OF LENGTH 5λ

these angles decrease with increasing antenna length. If the antenna is open at both ends and resonant, the net directional pattern extends equally in both directions. Where maximum radiation in only one direction is desired, the long-wire antenna may be terminated at one end in its characteristic impedance so as to be non-resonant. The radiation will then be largely concentrated in one direction, as indicated in Figure 31-12, where the driven end of the antenna is at the left and the terminated end at the right.

Combinations of long wires in appropriate geometrical arrangements are commonly used in long distance point-to-point practice for both transmitting and receiving antennas. One of the most effective such arrangements is the **rhombic antenna**
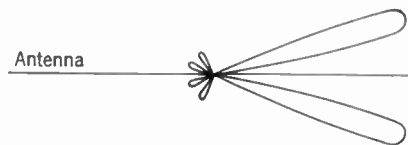
FIG. 31-12 RADIATION PATTERN OF NON-RESONANT LONG-WIRE ANTENNA

illustrated in Figure 31-13. This, as shown, consists of four wires arranged in a diamond or rhomboid pattern terminated in a resistance R which makes the system non-resonant. Each of the four wires is from two to four wavelengths long, and each sets up an individual radiation pattern, as indicated in the drawing. When the "tilt-angle" ϕ is properly chosen, four of the major radiation lobes point in the direction of the main axis of the rhombus while the other four major lobes tend in part to cancel each other out. The rhombic antenna operates satisfactorily over a wide frequency range because its impedance has a constant resistive value. This is of great importance in short-wave radio transmission over long distances where it is frequently necessary to change the frequency to obtain satisfactory transmission under changing radio propagation conditions. Rhombic antennas of identical design are used for both transmitting and receiving. A number of separate rhombics may be used, however, for "diversity reception" in which the energy received from several appropriately located antennas is combined to drive a single receiver.

For radio transmission in the very high and ultrahigh-frequency ranges, the basic form of antenna is a half-wavelength resonant wire commonly known as a **half-wave dipole**. The radiation pattern for this simple antenna, when isolated in
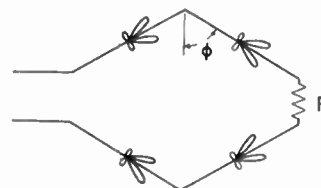
FIG. 31-13 RHOMBIC ANTENNA SHOWING RADIATION PATTERNS

space, is indicated in Figure 31-14. If the antenna is placed vertically, radiation strength is maximum along the horizontal and is equal in all directions. If the radiator is parallel to the ground, on the other hand, the field varies with azimuth and is zero in the direction of the wire axis. The vertical pattern is largely determined by the height of the antenna above the ground.

Increased directivity can be obtained by the use of arrays of half-wave antennas properly spaced and phased. One relatively simple arrangement employs one or more "parasitic antennas" in conjunction with a driven half-wave radiator. If maximum radiation in a particular direction is desired, a wire of such length as to be resonant at a frequency somewhat lower than the transmitted frequency, placed parallel to and at a distance
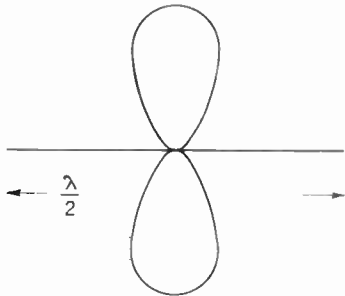
FIG. 31-14 RADIATION PATTERN OF ISOLATED HALF-WAVE ANTENNA

about 1/10 wavelength away from the driven antenna on the side opposite to the preferred direction of radiation, will act as a reflector. This parasitic reflector is not directly driven but the currents induced in it by the driven antenna set up a radiated field which combines with that of the driven antenna to produce a net field in the desired forward direction. Directivity may be further enhanced by using one or more additional parasitic wires in front of the driven antenna and spaced in tandem in the direction of transmission at separations of about .1λ. These wires are made of such length as to be resonant at a frequency higher than the transmitted frequency and act as "directors" to further concentrate the field in the desired direction.

When no directivity in the horizontal direction is desired, as in most broadcasting and mobile

radio applications, a vertical half-wave radiator may be employed to develop vertically polarized waves. This is the so called "flag-pole" type of antenna. Horizontally polarized omnidirectional
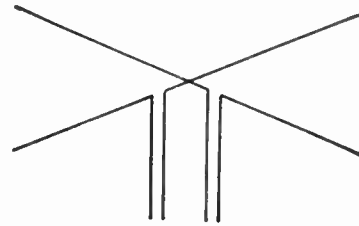
FIG. 31-15 TURNSTILE ANTENNA

waves are usually produced by antennas employing the "turnstile" principle. As indicated in Figure 31-15, the basic turnstile antenna consists simply of two half-wave resonant wire radiators placed crosswise at right angles to each other in the same plane. When the two radiators are excited by equal currents 90° out of phase with each other, the directional radiation pattern in the horizontal plane is roughly circular. Antennas of this type are advantageous in that they can be stacked vertically at half-wavelength intervals to increase horizontal radiation. They are used extensively with TV broadcast transmitters. The TV signal, however, is a very broad one (6 mc) and it is necessary that the antenna be able to operate over this broad frequency range. The broad-banding effect required is obtained by making the antenna itself thick so as to increase the effective diameter of the radiator. This may be accomplished by employing for the arms of the dipoles such structures as hollow cylinders, cones or cigar-shaped spheroids.

Radiation of electomagnetic energy in the superhigh-frequency range involves principles and methods that are analogous to, if not identical with, those of light transmission. Antennas used at these frequencies may be grouped under the general designation of aperture radiators. Their basic function is to transform the spherical wavefront, which is normally developed from a point source of radiation, to a plane wavefront. To the extent that this is accomplished, the radiated energy may be projected in a very narrow beam in the desired direction. In microwave practice, an-

Parabolic Reflector

Reflecting Element

Half-wave Dipole Antenna
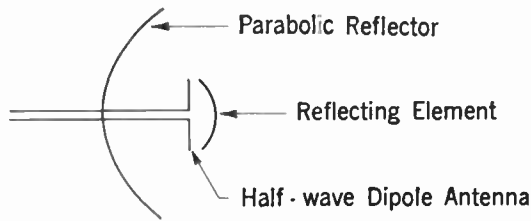
FIGURE 31-16

tenna gains in the order of 30 to 40 db are commonly obtained.

One method of obtaining a directive plane wavefront employs the optical technique of a parabolic reflector comparable to that of the ordinary searchlight. The geometrical characteristics of the parabola are such that waves emanating from a point source at the focus will be reflected in parallel straight lines that will all reach the plane of the mouth of the parabola at the same time. If, as illustrated in Figure 31-16, an auxiliary reflecting surface is placed in front of the energy source to prevent any direct radiation, all of the energy will be reflected from the paraboloid and will appear as a plane wave across its mouth. Since, as has already been pointed out, electromagnetic energy can escape directly from the open end of a waveguide, the same effect can be produced by means of a waveguide leading to the focal point of the parabola with its open end turned back to direct escaping energy toward the reflecting surface, as illustrated in Figure 31-17.

A plane wavefront across an aperture may also be attained by simply flaring out the end of a waveguide to form a long horn in which the fields in the guide can expand gradually to produce a

Rear Feed
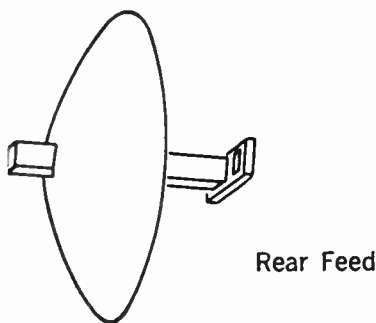
FIGURE 31-17

uniform field across the mouth of the horn. To accomplish this result, however, the flare angle of the horn must be small and an aperture comparable in size to that of the parabolic reflector could only be reached with a horn that would be too long to be practicable. One method of overcoming this difficulty, which has been applied extensively in the Bell System's TD-2 radio relay systems operating in the 4000 mc range, is the **delay lens antenna.** This consists of a comparatively short horn having a high flareangle but with a "lens" across its mouth that acts like an optical lens to produce a plane wavefront. This device is illustrated in Figure 31-18. As indicated there, the lens cross-section is of plano-convex form and is
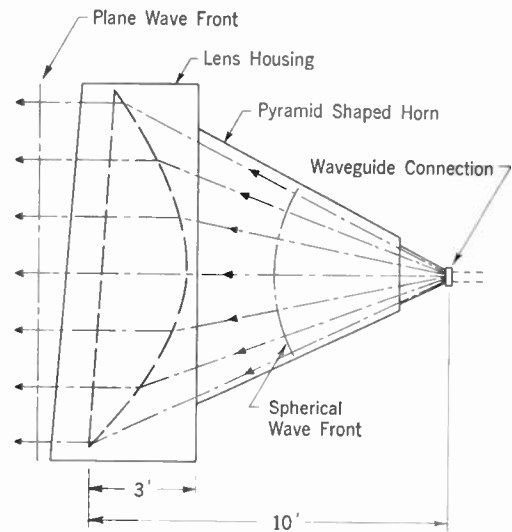
FIG. 31-18 DELAY LENS ANTENNA

placed at the mouth of a pyramid-shaped horn in a housing 10 feet square and 3 feet deep. The effect of the lens is much the same as that of a plano-convex lens on the transmission of light. Superhigh-frequency radio waves coming from the waveguide connected to the rear of the horn diverge, as indicated by the arrowed broken lines, to form a spherical wavefront. When they reach the lens, however, the velocity of the rays near the center is decreased enough by the thicker lens structure at that point to cause delay equal to the time required for the outer rays to traverse their geometrically longer path. All of the rays accordingly reach the front of the lens at the same time and proceed in the parallel paths of a plane wave-

[324]

front. The entire lens is tilted back slightly, as shown in the Figure. This is done to prevent any energy that might be reflected from the face of the lens from being focused back to the wave-guide outlet. Although its behavior is very similar to that of a glass lens with respect to light waves, this lens is made up of a large number of thin aluminum strips, each cut in the proper shape to form the overall plano-convex contour indicated in the drawing. The strips, which are mounted in slabs of foamed polystyrene, are parallel to each other and separated at a uniform, specified distance. The separation distance must have a certain relationship to the radio wavelength for the antenna to function, which means that a given design of this antenna is effective only over a limited frequency range. Also to be noted is the fact that the antenna will transmit only waves
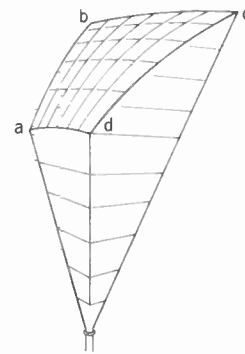
FIG. 31-19 HORN-REFLECTOR ANTENNA

RADIO RELAY STATION EQUIPPED WITH BOTH HORN-REFLEC-TOR AND DELAY-LENS ANTENNAS

that are polarized in the same direction as the parallel metal strips. Within these limitations, however, it has a gain of up to 45 db.

Somewhat more versatile is the **horn-reflector antenna** illustrated in Figure 31-19 and the accompanying photograph. This combines the virtues of the horn and the parabolic reflector. In this case, however, the reflecting surface, abcd, is only a small section of an imaginary parabolic "dish" some forty feet in diameter. The apex of the horn is at the focal point of this imaginary dish. Waves emanating from the circular wave-guide connected at that point are confined within the straight walls of the horn. The parabolic reflecting surface, which must be accurately shaped to one-sixteenth of the shortest wavelength, is so placed as to turn the direction of transmission through an angle of 90°. It also converts the spherical wave fronts coming up the horn into plane wave fronts at the 64.5 square foot aper-

FIG. 31-20 TD-2 TRANSMITTING TERMINAL

ture, which results in a beam concentrated within about 2°. This antenna has a gain of about 40 db at TD-2 frequencies (4000 mc) and somewhat higher gains at shorter wavelengths. It also transmits with equal facility waves polarized in any direction.

31.5 Microwave Radio Relay Systems

A number of types of radio relay systems employing frequencies in the SHF range are in use in long distance telephone work. Most extensively used over long-haul routes up to the present time is the system coded TD-2, which operates in the frequency range between 3700 and 4200 mc. In the usual application, the 500 mc band is divided into 12 channels, each with a bandwidth of 20 mc and separated from adjacent channels by 40 mc. Each 20 mc channel can handle one one-way television signal or one direction of transmission of the many voice channels of a Type-L telephone carrier system. All transmitting channels are served by one antenna, and all receiving channels are served by a second antenna of identical design. The system therefore requires only two antennas at terminals and four at relay points. In practice, two 20-mc channels are used for each of six two-way channels numbered as indicated below:

Channel No.	Frequencies
1	3720-3740 mc and 3760-3780 mc
2	3800-3820 mc and 3840-3860 mc
3	3880-3900 mc and 3920-3940 mc
4	3960-3980 mc and 4000-4020 mc
5	4040-4060 mc and 4080-4100 mc
6	4120-4140 mc and 4160-4180 mc

At each relay or repeater station, the channel frequencies are interchanged so that the outgoing signal is always 40 mc removed from the incoming signal. This shifting of the channel frequencies avoids possible difficulties that might be caused by local feedback between transmitting and receiving antennas.

Figure 31-20 is a simplified block schematic of the TD-2 transmitting terminal. As there indicated, the input signal, which may cover all or part of a band between 30 cycles and about 4 megacycles, is first amplified by a video amplifier, which increases the amplitude of the signal waves to a maximum

of 8 volts peak-to-peak. When multiplex telephone signals are being transmitted, the output of the video amplifier is applied directly to the FM modulator. For television, a clamping circuit is included between the amplifier and modulator. This adds a d-c component to the signal to clamp it to the amplitude value of the tips of the horizontal synchronizing pulses as a base line.

Frequency modulation is obtained in a comparatively simple manner by applying the signal directly to the repeller electrode of a reflex-klystron oscillator, which is tuned to a nominal frequency of 4280 mc. A signal level of 8 volts peak-to-peak at the repeller produces ± 4 mc frequency modulation. The frequency-modulated carrier is then applied to a converter, whose active component is a point-contact varistor, where it is mixed with the output of a 4210-mc beating oscillator. The resultant converter output is an intermediate frequency normally centered about 70 mc and ranging between 74 and 66 mc. This of course represents the difference between the two frequencies applied to the converter. The four-stage IF amplifier raises the level of the modulated IF signal to +13 dbm. A small portion of the amplifier output is picked off and applied to a slow acting automatic frequency control circuit, which, in the case of telephone transmission, measures the average frequency at the amplifier output and adjusts the beating oscillator so that this average frequency remains constant at 70 mc. For video transmission, the AFC circuit measures the amplifier output frequency only during the horizontal synchronizing pulses and adjusts the beating oscillator frequency to hold this frequency constant at 74 mc.

MICROWAVE TRANSMITTER-MODULATOR

Microwave Receiver

F M Receiving Terminal

Channel
Filter

Image
Rej
Filter

Rec
Con-
verter

IF
Pre - Amp

Main
IF
Amp

Through
Patching
Circuits

Limiter

Discrim-
inator

Output
Signal

Video
Amp

AVC

Micro-
wave
Gen

To 5 Other
Channel Filters

⊟— Indicates Waveguide
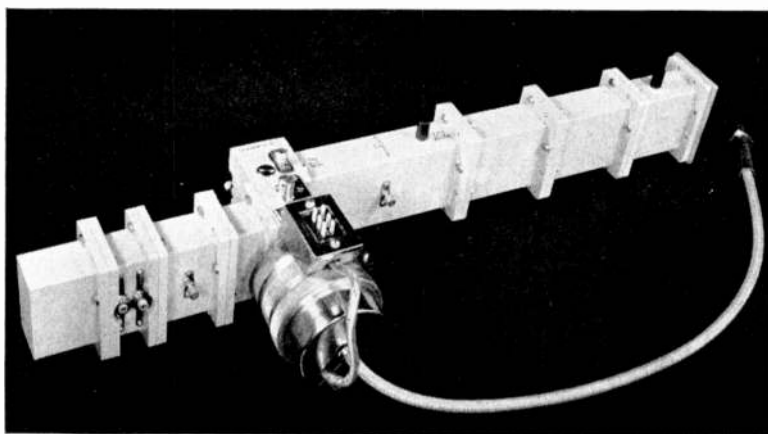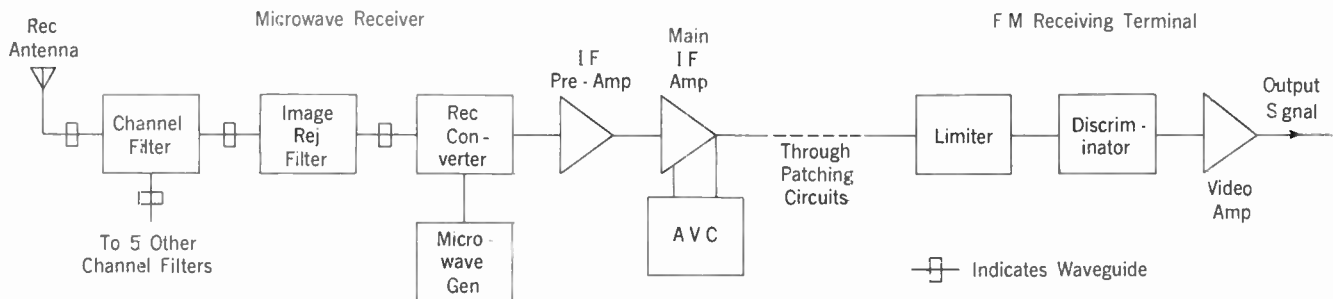
FIG. 31-21 TD-2 RECEIVING TERMINAL

All the components discussed above comprise what is generally known as the FM transmitting terminal. The modulated signal output of this terminal, after passing through appropriate switching or patching circuits, is applied to the microwave transmitter itself. The transmitter modulator translates the IF signal to the desired microwave frequency. The modulator employs a 416 type tube, and is supplied with the proper beat-frequency from a microwave generator whose output frequency is 70 mc removed from the desired microwave channel frequency. The basic unit of the microwave generator is a very stable crystal-controlled oscillator operating in the frequency range from 17.5 to 19.0 mc, depending on the frequency of the crystal employed. The basic oscillator is followed by a series of frequency-multiplying stages providing a total multiplication factor of 216. The modulated output is led through a band-pass waveguide filter, which selects the upper side-band. The microwave transmitter amplifier, which was described in Article 25.3, is capable of producing an output of +27 dbm (slightly more than $\frac{1}{2}$ watt). This output is fed through a channel filter where it is joined by the outputs of five other transmitting channels and applied collectively to the transmitting antenna.

At the receiving terminal of a microwave channel, the incoming channels are separated by channel filters, as indicated in Figure 31-21. The incoming energy for each channel then passes through an image rejection filter. This is a bandpass filter designed to have particularly high suppression characteristics in the neighborhood of the image frequency 140 mc away from the signal frequency. The receiver converter is a demodulator employing point-contact varistors, in which the SHF signal is mixed with the output of a microwave generator to again produce the 70 mc IF. This is passed through an IF pre-amplifier having a gain of about 12 db to an 8-stage main IF amplifier having a maximum overall gain of about 60 db. Associated with this amplifier is an automatic volume control circuit which compensates for differing input levels due to fading and holds the output power constant at approximately +9 dbm. The FM receiving terminal includes limiting and discriminating circuits which convert the frequency-modulated 70 mc signal back to its original amplitude-varying form in the frequency range between 30 cycles and 4 megacycles. This signal is applied to a video amplifier whose push-pull output voltage is about 2.0 peak-to-peak.

Repeater stations, which are located at intervals of about 25 miles along the radio relay routes, are of two types—main and auxiliary. Main stations include switching and branching circuits, while auxiliary stations are arranged only to receive, amplify and re-transmit the radio signal. The layout of the main repeater station may be represented by connecting together, through patching circuits, a microwave receiver as shown schematically at the left of Figure 31-21 and a microwave transmitter as shown at the right of Figure 31-20. The 40-mc frequency shift that is made at all repeaters is obtained by using conversion frequencies that differ by 40 mc. In the auxiliary repeater arrangement which is shown in block schematic in Figure 31-22, a single microwave generator is used to supply both the receiver converter and the transmitter modulator. The 40-mc shift is secured by the use of a "shifter converter" which mixes the microwave generator output with the output of a separate 40-mc oscillator to provide a supply for the receiver converter

that differs by 40 mc from the microwave generator frequency.

It will have been noted that the general operating principles of this microwave system have much in common with the more usual radio and carrier systems. Because of the extremely high frequencies employed, however, most of the apparatus units differ radically in design from those of relatively low-frequency systems. One of the
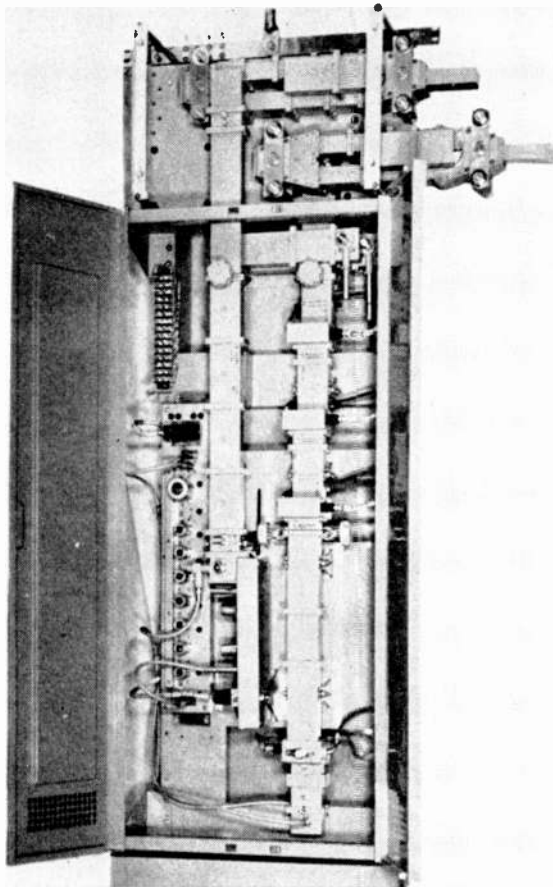
most interesting examples of such design difference is found in the channel filters of the TD-2 system. The key component of these filters is a device known as a waveguide hybrid, one form of which is illustrated in Figure 31-23 together with its circuit analog. When the impedances of the four waveguide arms are properly matched, en-

ergy entering arm C will divide equally between arms A and B and none will reach arm D. Similarly, energy applied at D will divide equally between arms A and B with no output to C. However, when the input is to arm C, the outputs of arms A and B are in phase opposition, while with the input at D, the outputs at A and B are in phase. This may be understood by referring to the circuit analog. It follows that when equal and in-phase voltages are applied across arms A and B there will be no output to arm D and full output to arm C. On the other hand, if the equal inputs of arms A and B are 180° out of phase, there will be no output at C and full output at D.

The arrangement for employing these waveguide hybrids to obtain filter action is illustrated schematically in Figure 31-24. The total microwave energy coming from the antenna enters the upper hybrid at arm C and divides equally between A and B with no transmission at D. Inserted in series with both arms A and B are identical band-reflection filters (Fig. 20-22) tuned to reflect the frequency band of one particular microwave channel but to pass all other frequencies. The frequencies of the reflected channel band travel back to the hybrid and are applied to arms A and B. Because one of the band-reflection filters is located one-quarter wavelength farther away from the hybrid than the other, the energy reflected by one has to travel a half-wavelength farther than that traveled by the other in going from the hybrid to the reflection filter and back. The reflected waves are therefore 180° out of phase when they reach the hybrid and the total reflected energy is therefore transmitted to arm D and thence to the channel receiver.

The energy that was not reflected by the filters passes on to arms A and B of the second hybrid in phase and is accordingly transmitted to arm C. This is connected to another arrangement identical to that of Figure 31-24 except that its filters are tuned to reflect a different channel band; and so on until all channels have been dropped off to their respective channel receiving circuits.

The TD-2 system, as discussed above, provides a separation of 40 mc between adjacent 20 mc channels in order to minimize the possibility of interference between channels. By employing horn-reflector antennas and polarizing adjacent channels alternately in vertical and horizontal directions, it is possible on most routes to add an additional channel in the space between each two original channels and thus increase the total num-
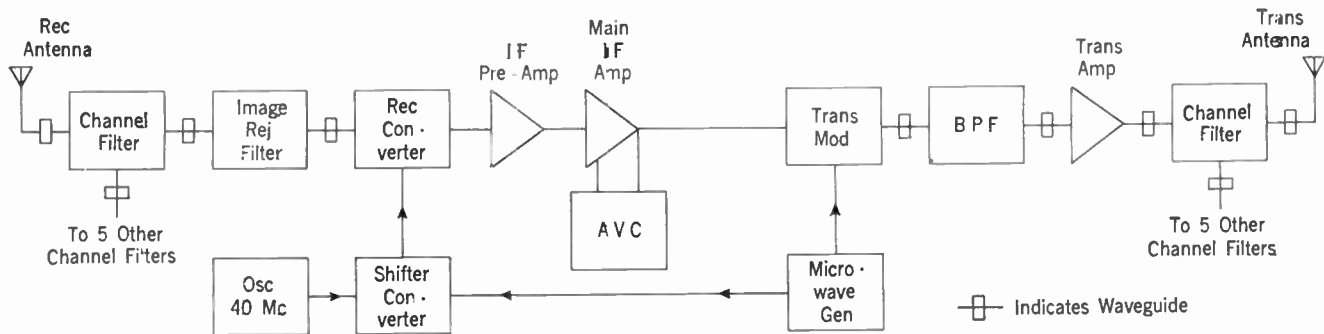
FIG. 31-22 TD-2 AUXILIARY REPEATER

ber of 20 mc channels on the route to 24. This reduces the separation between adjacent channels from 40 to 20 mc but interference between channels is avoided by the cross-polarization and by the addition of special filters in the IF circuits of all channels at both terminal and repeater points.

Another microwave radio system, coded TH in the Bell System, is designed to operate over the same routes as the TD-2 systems and to share the same horn reflector antennas and circular waveguides. The TH system operates in a 500 mc band extending between 5925 and 6425 mc and provides a maximum of eight broadband channels in each direction. Each such channel accommodates a 10 mc baseband signal which can carry a maximum of 1860 telephone message channels, or a 4.5 mc TV signal plus as many as 420 voice channels.

The broadband channels are divided into two groups with eight transmitting channels occupying one side of the 10 mc baseband and eight receiving channels the other side. Within each group, the channels are spaced 29.65 mc apart. A guard band is left between the two groups so that there is always a 250 mc difference between the transmitting and receiving sides of each two-way channel. In each group, too, the channels are alternately polarized vertically and horizontally so that the minimum separation between channels with the same polarization is 59.3 mc.

When a TH channel is employed exclusively for telephone message circuits, three 600 voice-channel master groups and one 60 channel supergroup are allocated to the 10 mc baseband in the same frequency ranges as used in L3 coaxial cable systems. For combined TV and telephone message operation, the 4.5 mc TV signal is applied to the

lower part of the 10 mc baseband. The 420 message channels are included in the range of master group 3 between 5.764 and 8.284 mc. It is necessary in this case, however, to leave an unused gap about 800 kc wide in the neighborhood of 7.2 mc in order to avoid possible interference from the second harmonic of the 3.58 mc television color subcarrier signal.

As might naturally be expected, the basic radio circuit elements of the TH system are similar in principle to those of the TD-2 system. There are, however, a number of important differences in the circuitry and in the kinds of equipment used. As in the case of the TD-2 transmitting terminal shown in Figure 31-20, each TH channel employs an FM terminal and a microwave radio transmitter at the sending end. The 10 mc baseband signal, after amplification in an appropriate video amplifier, is applied to the repeller electrode of a
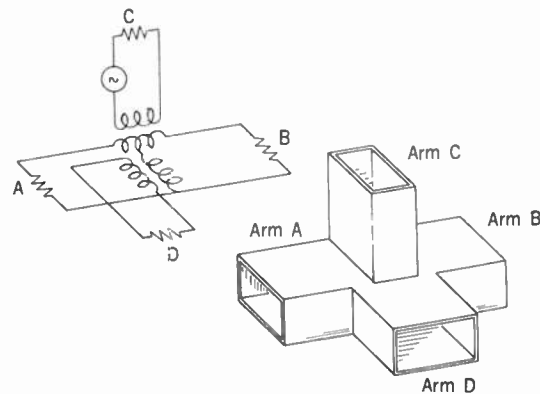
FIG. 31-23 WAVEGUIDE HYBRID AND CIRCUIT ANALOG
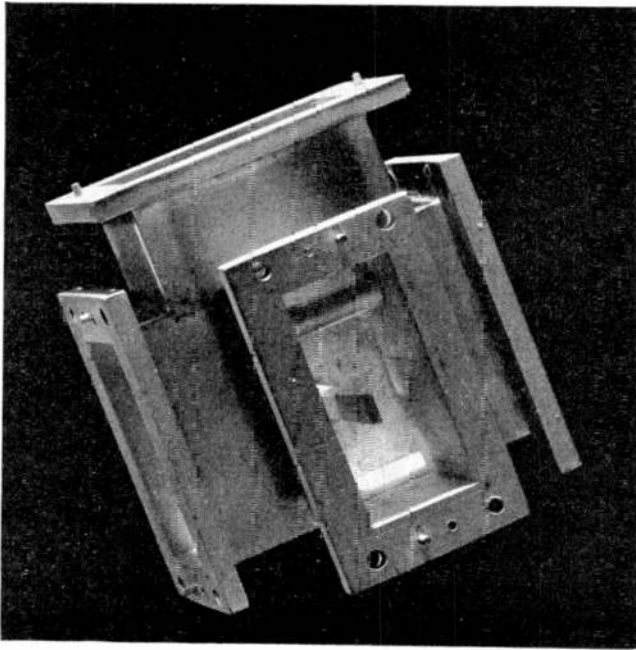
ONE FORM OF WAVEGUIDE HYBRID

reflex-klystron oscillating at a nominal frequency of 6000 mc. The frequency-modulated output is fed to a converter along with the output of a separate klystron beating oscillator whose frequency is maintained at a difference of 74.1 mc from that of the modulating oscillator by means of an automatic frequency control circuit. The output of the converter is thus a 74.1 mc frequency-modulated IF carrier. This is amplified and supplied, through appropriate patching connections, to the microwave radio transmitter.

The major elements of the radio transmitter circuit are indicated in the block diagram of Figure 31-25. The frequency-modulated IF signal is first passed through a three-stage limiting circuit which removes any amplitude modulation that may be present and also adds some amplification. The IF signal is then mixed in the transmitter modulator with the appropriate carrier frequency to produce the desired microwave signal for the particular channel. As shown in the drawing, the carrier frequency is obtained from a carrier supply modulator in which either of two microwave frequencies may be mixed with either of two VHF frequencies. These four basic frequencies—29.65, 59.3, 6049, and 6301 mc—are developed in a microwave carrier supply circuit, common to the whole office, in which the output of a stabilized

oscillator operating at 14.826 mc is multiplied by 2,4,408 and 425, respectively. Proper combinations of these basic frequencies in the carrier supply modulator will produce beat frequencies which will develop any one of the desired 16 channel frequencies at the output of the transmitter modulator. The channel frequency of 6034.2 mc, for example, is obtained by combining 6049 and 59.3 in the carrier supply modulator to get the sum output of 6108.3 mc. This beating frequency is then combined with 74.1 mc in the transmitter modulator to produce the desired 6034.2 mc as the difference frequency. The output of both modulators of course includes both sum and difference frequencies but the band pass filter which follows the modulator in each case permits transmission of only the desired sideband.

The carrier frequency supplied by the carrier supply modulator and its associated band-pass filter is amplified to the high level required for proper operation of the transmitter modulator by a traveling-wave tube amplifier. Its output is passed through another filter and a ferrite isolator in order to suppress respectively any harmonics that might be generated in the traveling-wave tube and any reflections that might be returned from the transmitter modulator. The band-pass filter following the transmitter modulator passes the desired sideband of the frequency-modulated microwave signal and reflects the other sideband back to the ferrite isolator where it is absorbed. The signal is then amplified by the
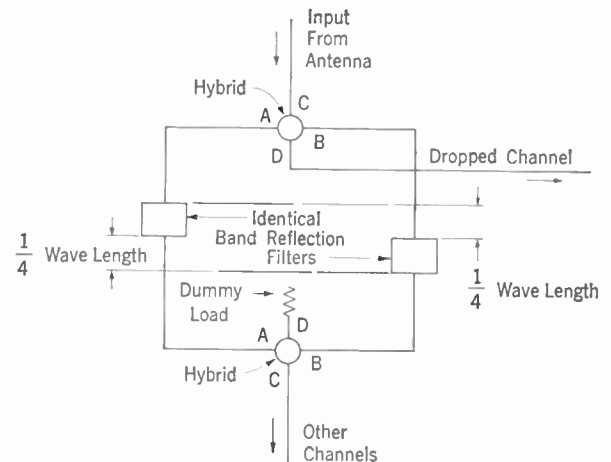
FIG. 31-24 CHANNEL BRANCHING FILTER PRINCIPLE
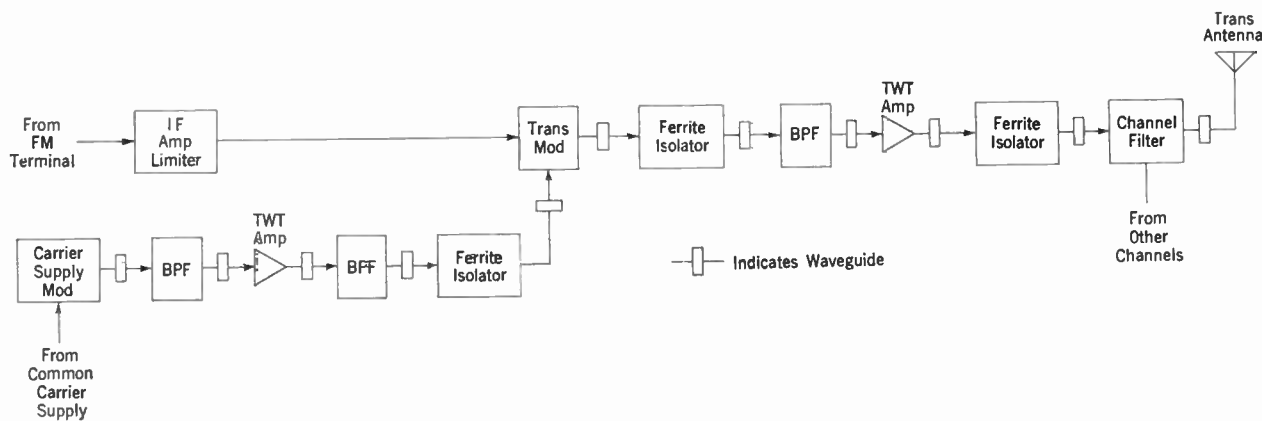
FIG. 31-25 TH RADIO TRANSMITTER

following TWT amplifier to a level of $+37$ dbm (5 watts) and transmitted to the channel filter network where it is combined with other TH channels and carried by circular waveguide to the transmitting antenna. Another ferrite isolator is inserted between the amplifier and the channel filters to suppress any reflections from the immediately following system combining networks or the antenna waveguide.

The TH receiving terminal circuits are fundamentally the same as those of the TD-2 system illustrated in Figure 31-21. The microwave beating frequencies supplied to the receiver converter or modulator are obtained from a carrier supply modulator in the same way as described above for the transmitting terminal except that no amplifier is required in this case. Ferrite isolators are inserted in front of the converter in both the carrier supply circuit and the main channel circuit to absorb any reflections from the modulator. As in the TD-2 system, the IF signal at the output of the modulator is passed through a low-noise pre-amplifier and a main amplifier with automatic gain control, and hence through appropriate patching circuits to the FM receiving terminal. This includes an amplifier-limiter, discriminator and video amplifier as in Figure 31-21.

A third microwave radio system, coded TJ, is used in the Bell System for relatively short-haul applications not exceeding about 300 miles. This operates in the frequency range between 10,700 and 11,700 mc and provides a maximum of six two-way broadband channels. The total range of 1000 mc is divided into twenty-four channels, each about 40 mc wide. Twelve of these channels are used in the section or "hop" on one side of a re-

peater and the alternate twelve channels for the hop on the other side of the repeater. The spacing between adjacent channels in any repeater section is thus 80 mc. The twelve channels for each section are divided into two groups of six, with the higher group transmitting and the lower group receiving at one repeater station and vice versa at the next station. Channels are polarized alternately between vertical and horizontal so that there is a minimum separation of 160 mc between channels of like polarization.

The relatively simple circuit arrangement at terminal and repeater stations is shown schematically in Figure 31-26. For transmitting, the baseband signal, after amplification, is applied to the repeller electrode of a reflex-klystron whose basic frequency is automatically controlled by means of a microwave discriminator circuit. The frequency-modulated output of the oscillator, at a power level of about 0.5 watt, is supplied directly to the channel combining filter without further amplification. A ferrite isolator is included in its path, however, to suppress microwave reflections from the filters or antenna. For receiving, the selected microwave channel band is shifted to a 70 mc IF signal in a conventional converter circuit by heterodyning with a controlled beating frequency supplied by a klystron oscillator. The IF signal, after amplification, is demodulated to baseband by a limiter-discriminator circuit, as indicated. In the TJ system, the microwave signals are converted to baseband at intermediate repeater points as well as at the terminals. This is in contrast with the TD-2 and TH systems where the signals may be transmitted through intermediate repeaters without frequency demodulation.

Where routes coincide, TJ system signals may be transmitted and received simultaneously with TD-2 and TH signals through the same circular waveguides and horn-reflector antennas. Where, as may most commonly be the case, a TJ route is not shared with other systems, it is generally more economical to employ a "periscope" type arrangement. In this a small paraboloidal dish-type antenna is mounted horizontally on the roof of the building in which the terminal or repeater equipment is housed. This directs a narrow beam vertically up to the top of a tower on which a plane or curved reflector is mounted in such a position as to redirect the radio beam horizontally in the desired direction. Such an arrangement minimizes waveguide losses while still providing good antenna gain.

In all three of the microwave radio systems discussed above, the channel combining and selecting filters employ waveguide hybrids and band reflection filters operating in the manner described in connection with Figure 31-24. When more than one system, or more than one polarization, is operated on the same antenna, it is necessary to provide additional filtering devices to combine or separate the several systems and polarizations. An assembly of waveguide components, known collectively as a "systems combining network" is employed for this purpose.

If all three systems are used, and each employs two polarizations, there are six classes of signals to be separated one from the other. The 2.81 inch (inside diameter) circular waveguide which connects the radio equipment to the horn-reflector antenna is capable of simultaneously carrying all of these signals in their dominant TE_{11} modes. Such a circular waveguide, however, is also capable of supporting a number of other modes in each of the three frequency ranges. In order to avoid energy losses and reflections due to mode conversions, therefore, it is necessary that the guide be free from any irregularities or discontinuities. This means that the guide must run straight down the tower and that the systems combining network must be placed in the same vertical run at the foot of the tower. In the manufacturing process, every effort is made to insure that the guides are truly circular in cross-section at all points. However, all but very short guides are installed with a tuning section, known as an "axial ratio compensator" at the lower end. This is a twelve inch section of circular guide which has been purposely deformed to be somewhat out-of-round. At the time of installation, it may be rotated into a position that tends to counteract the effects of slight amounts of ellipticity in the guide proper.

The first component of the systems combining network, which is connected below the guide tuner, is a transducer which tapers the circular guide down to a 1.79 inch square guide, a size sufficient to support transmission of all of the frequencies concerned. The transducer is followed by two coupling networks in tandem, one for each of two 4000 mc polarizations. Following these, another transducer tapers the 1.79 inch square guide down to a 1.27 inch square guide. The latter supports transmission in the 6000 and 11,000 mc ranges but rejects 4000 mc. The coupling network mentioned above consists of a section of rectangular waveguide, of the proper dimensions to support the 4000 mc dominant mode, running parallel to and with its narrow face adjacent to the 1.79 inch square guide. A longitudinal slot cut in the common wall couples to one polarization of the 4000 mc signal. The rectangular guide is
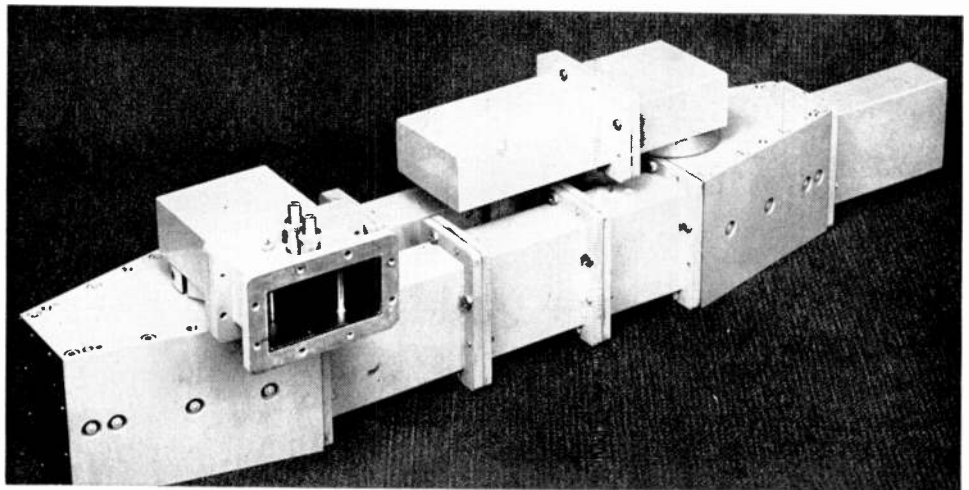
CHANNEL SEPARATION NETWORK
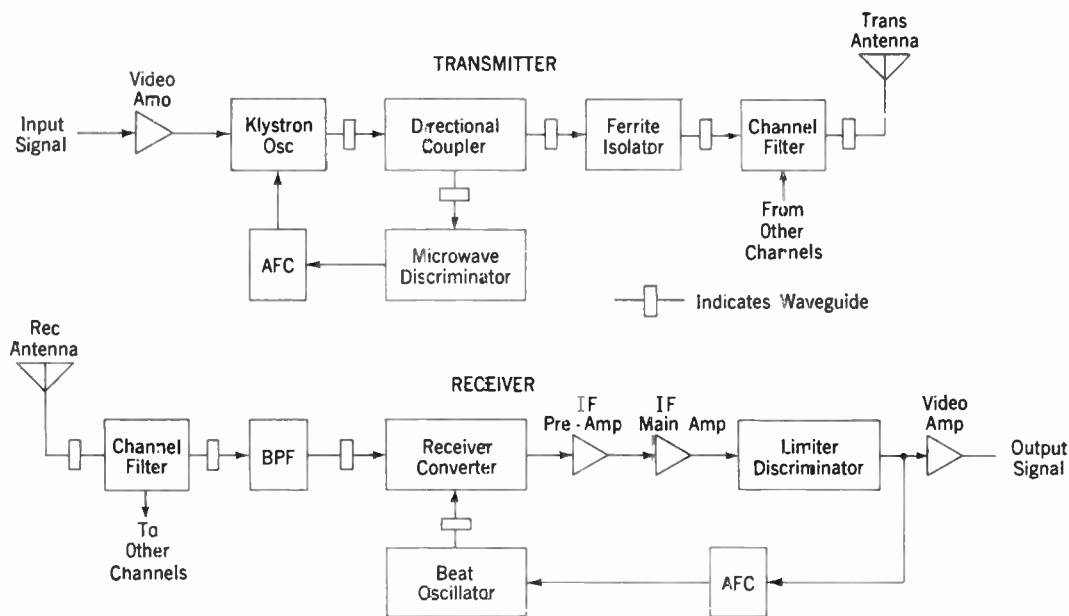
FIG. 31-26 TJ RADIO TERMINAL

then extended to the appropriate channel filter network. Since this guide can transmit only the desired dominant mode, it may include such reasonable bends or twists as may be necessary to lead it into the building housing the channel equipment. The second coupling network is exactly the same except that the rectangular guide section is rotated 90° with respect to the square section to couple to the other 4000 mc polarization.

Two similar coupling networks are connected below the 1.79 to 1.27 inch transducer to extract the two 6000 mc polarizations. These are followed by another transducer which tapers from 1.27 inch square to .75 inch circular guide. The latter supports 11,000 mc waves but rejects the lower frequencies. One 11,000 mc polarization is taken

off in a longitudinal coupler as above. The other polarization is extracted at the lower end of this coupling section through another transducer which tapers the .75 inch circular guide to a rectangular guide of proper dimensions for handling the dominant mode at 11,000 mc.

When all three frequency ranges in both polarizations are involved, the complete systems combining network, as described above has a total length of about fifteen and one-half feet. In many installations. however, not all of the six possible signal classes are involved. Certain components of the combining network may then be omitted accordingly, with a consequent reduction in its overall length.

CHAPTER 32

NOISE AND CROSSTALK

32.1 Induced Effects in Telephone Circuits

One of the factors upon which the intelligibility of a telephone conversation depends is the absence of excessive noise and crosstalk. If each telephone circuit was completely isolated from all other telephone circuits or other electric circuits of whatever kind, including earth currents and atmospheric charges, we would not expect any potentials to exist in the telephone circuit other than those deliberately introduced for the purpose of the desired signal transmission. However, this is in fact a purely hypothetical situation as, in practice, nearly every long telephone circuit is in close proximity to other telephone circuits, and sometimes to other electric circuits such as power lines. It is necessary, therefore, that telephone circuits not only be efficient in transmitting electric energy without distortion and without too great a loss, but also that they be protected against induced voltages caused by adjacent telephone circuits or other electric circuits.

As a matter of fact, any two long paralleling telephone circuits that are not "balanced" against each other by means of transpositions, or otherwise, will crosstalk to such a degree as seriously to interfere with their practical use. Furthermore, because crosstalk is largely an inductive effect, its magnitude tends to increase with (1) the length of the paralleling circuits, (2) the strength (energy level) of the transmitted currents, and (3) the frequency of the transmitted currents. It follows that the use of telephone repeaters is likely to increase the crosstalk possibilities because these devices permit longer circuits and at the same time increase the level of the energy at certain points along the line. Crosstalk possibilities are also increased by the use of carrier systems because of the higher frequencies employed.

Of course, crosstalk can be caused by the direct leakage of current from a disturbing to a disturbed circuit. With properly maintained lines, however, insulation is usually sufficiently good to make this a negligible factor. The crosstalk coupling which presents the real problem in prac-

tice is due to the electric and magnetic fields set up by the currents in the disturbing circuit. The effects of these two fields are not entirely alike, although their results are generally similar as far as crosstalk is concerned.

32.2 Causes of Crosstalk

As we know, the effect of the magnetic field of one circuit on a second paralleling circuit is called magnetic induction. Similarly, the effect on the second circuit of the electric field of the first circuit is called electric or electrostatic induction. How magnetic induction causes crosstalk can be seen by referring to Figure 32-1. This shows the four wires of two telephone circuits running parallel to each other in the usual flat configuration employed on standard open wire lines. If we assume that an alternating voltage is applied to the left end of the disturbing circuit, A, which is made up of wires 1 and 2, the magnetic field existing about a short section S of this circuit may be represented by lines of magnetic induction as shown.

At a particular instant in the alternating cycle, the current in wires 1 and 2 may be represented

FIGURE 32-1

[334]
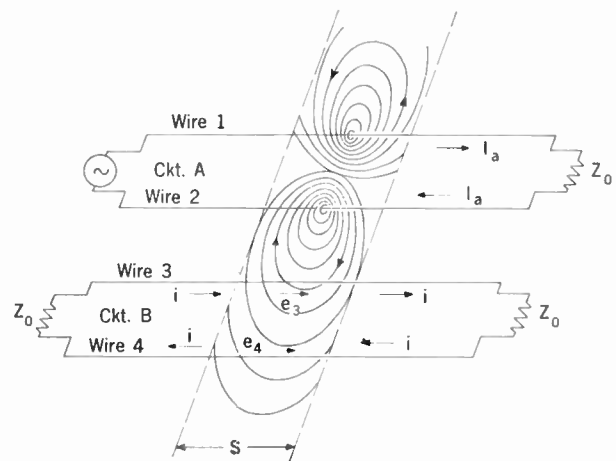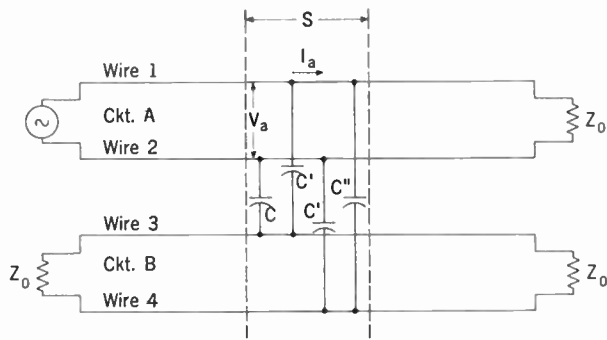
FIGURE 32-2

by the equal and opposite vectors designated I_a. As I_a increases or decreases in value, the associated lines of magnetic induction will cut wires 3 and 4 of the paralleling telephone circuit, B. But with the relative spacing of the wires shown in the diagram, more lines will cut wire 3 than cut wire 4. Accordingly, the voltage induced by the magnetic field in wire 3 will be somewhat greater than that induced in wire 4. The voltages induced in both wires are in the same direction at any given instant, so that they tend to make currents circulate in circuit B in opposite directions. If they were equal, therefore, their net effect would be zero. But in so far as the induced voltage e_3 exceeds the induced voltage e_4, there is an unbalance voltage, $e_3 - e_4$, tending to make a current circulate in circuit B. If the circuit is terminated at both ends in its characteristic impedance, Z_0, the current resulting from this unbalance voltage induced in a short section of the circuit may be written as—

$$i = \frac{e_3 - e_4}{2Z_0} \qquad (32:1)$$

The summation of the circulating currents caused by the voltages induced in each small section of the line may result in an appreciable current through both terminals of the circuit, which is due entirely to magnetic induction.

It should be noted that although the energy in the disturbing circuit was considered in the above as being transmitted from the left end to the right end of the circuit, the crosstalk current in the disturbed circuit appears at both ends of the circuit. The crosstalk appearing at the left end of the disturbed circuit is known as **near-end crosstalk** and that appearing at the right end is known as **far-end crosstalk.**

Now turning our attention to electric induction, Figure 32-1 may also be used to show the equipotential lines of the electric field established about circuit A under the same conditions as in the previous example. This electric field will set up potentials on the surfaces of wires 3 and 4 and, with the configuration and spacing of the wires shown in the drawing, these potentials will not be equal. The resultant difference in potential between wires 3 and 4 will tend to cause crosstalk currents to flow to both ends of circuit B.

The crosstalk effect of electric induction may also be analyzed by consideration of the capacitive relationships between the wires of the disturbing and disturbed circuits. Thus, referring to Figure 32-2, we know that in any unit length of the two circuits there is a definite capacitance between wire 1 and wire 2 and between wire 3 and wire 4. Moreover, if the wires are equally spaced as shown, the separation between wires 2–3 is the same as that between wires 1–2 or 3–4, and there is therefore the same capacitance between wires 2–3 as between the wires of either pair. This is represented in the Figure by the small capacitor designated C. Similarly, the capacitances between wire 1 and wire 3 and between wire 2 and wire 4 are designated by capacitors, C', less in value than C because the separation between these wires is greater. There remains also the still smaller capacitance between wire 1 and wire 4, which is indicated by C''. Now if we assume an alternating current flowing in circuit A, there will be a difference of potential between wires 1 and 2, which will tend to cause small currents to flow through the several capacitances to the wires of circuit B. The net effect can best be analyzed by redrawing the diagram of Figure 32-2 in the form of a Wheatstone bridge network as shown in Figure 32-3. A study of the capacitance values of the arms of this bridge shows that the impedances of the arms are not such as to give a balanced con-
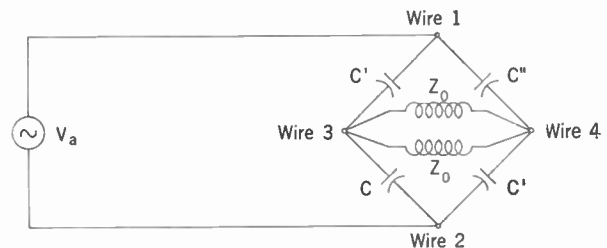
FIGURE 32-3
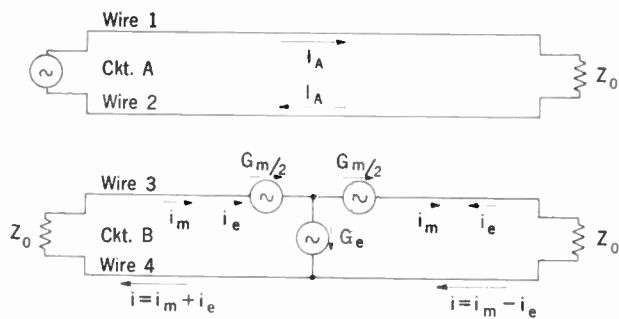
FIGURE 32-4

dition and, consequently, current flows through the impedances Z_0. In other words, a current is set up in circuit B which will manifest itself as crosstalk at both ends of the circuit.

As shown in Figure 32-4, the crosstalk due to electric induction may be thought of as being caused by a small generator G_e connected across the disturbed pair, while the crosstalk due to magnetic induction may be thought of as being caused by a generator G_m connected in series with the disturbed pair. Under these conditions, it will be noted that the currents established by the two generators flow in the same direction in the left portion of the line, but in opposite directions in the right portion. In other words, the crosstalk effects of magnetic and electric induction are additive in the case of near-end crosstalk, but opposed to each other in the case of far-end crosstalk.

32.3 Principles of Crosstalk Reduction

There are a number of possible ways of eliminating, or at least substantially reducing, the crosstalk induction discussed in the preceding Article. One possibility is to arrange the paralleling wires in such a configuration that the effect of the field of one pair will be the same at both wires of the
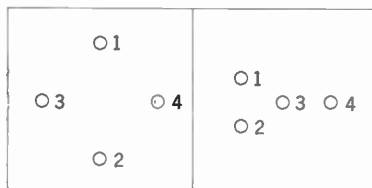
FIGURE 32-5

other pair, thus leaving no residual difference to cause currents in the second pair. Two possible ways to effect such a non-inductive configuration are shown in Figure 32-5. For a number of reasons, however, such wire configurations are not usually practicable.

Another partial solution is to reduce the separation between the wires of either or both disturbing and disturbed pairs and, if practicable, at the same time to increase the separation between the two pairs themselves. A glance at Figure 32-1 will show that if the two wires of the disturbing pair are spaced closely together, the fields set up by the two wires will occupy approximately the same position and will therefore tend to neutralize each other. Similarly, if the two wires of the disturbed pair are close together, the effect of any
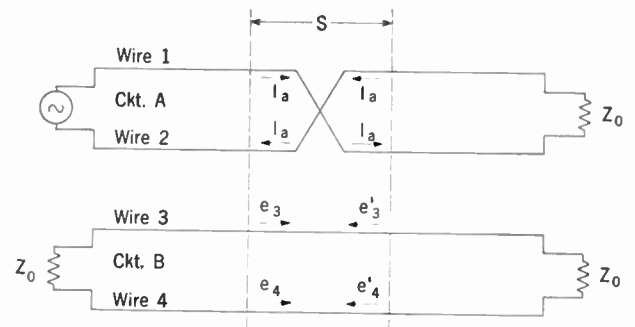
FIGURE 32-6

field set up by the disturbing pair will be practically the same on both wires of the disturbed pair, so that there will be no resultant unbalance voltages to produce crosstalk. However, it is possible to take advantage of these factors in practice to only a limited extent. In cable circuits, the two wires of each pair are close together, but so are the pairs themselves; in open wire circuits, there is considerable separation between pairs, but the two wires of a pair cannot be placed close together.

A third alternative is the use of transpositions. The principle involved here can be understood by referring to Figure 32-6, which shows the same four wires as were indicated in Figure 32-1. In this case, however, the relative pin positions of wires 1 and 2 are interchanged by means of a transposition in the middle of the short section, S. Under these conditions, it will be evident that when equal and opposite currents are flowing in
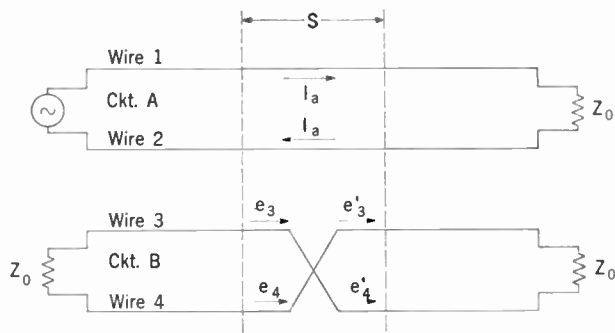
FIGURE 32-7

the two wires of circuit A, the voltages induced in wires 3 and 4 will be in opposite directions on the two sides of the point where circuit A is transposed. Thus, as indicated in the Figure, while e_3 is larger than e_4 and e'_3 is larger than e'_4, e_3 is exactly equal and opposite to e'_3 and e_4 is exactly equal and opposite to e'_4. There is therefore no net voltage induced in either wire 3 or wire 4 and, consequently, no crosstalk from circuit A.

The same net effect would be obtained by inserting the transposition in the disturbed circuit B and leaving the wires of the disturbing circuit running straight through, as shown in Figure 32-7. In this case, the voltage induced in the wire nearer wire number 2 is broken into two equal parts represented by the vectors e_3 and e'_3. Similarly, the voltage induced in the wire farther away from wire number 2 is broken into two parts, e_4 and e'_4. But with the transposition as shown, voltage e_3 combines with voltage e'_4 and voltage e_4 combines with voltage e'_3. The induced or crosstalk current in the section, therefore, is—

$$i = \frac{(e_3 + e'_4) - (e_4 + e'_3)}{2Z_0} \qquad (32:2)$$

But with the transposition in the center of the section as shown, it is obvious that—

$$e_3 + e'_4 = e_4 + e'_3$$

The numerator of equation (32:2) is therefore equal to zero and there is no resultant crosstalk.

Either of the above discussed transpositions would be equally effective in reducing crosstalk due to either magnetic or electric induction. It may be noted, however, that a transposition at the same point in both circuits would have no such effect.

While a single transposition as discussed above is effective in eliminating crosstalk in a short section S, it would not be sufficient to reduce cross-

talk in the whole line. There are two principal reasons for this. First, because of attenuation, the current and voltage near the energized end of the disturbing circuit are many times as great as the current and voltage near the other end of the circuit. It could not be expected, therefore, that the induced crosstalk on the energized side of the transposition would be neutralized by the weaker crosstalk induced on the other side of the transposition. As a matter of fact, even in a short section, the transposition will not completely eliminate **near-end** crosstalk because the currents coming back from the far side of the transposition are necessarily attenuated somewhat more than the currents coming back from the near side of the transposition. On the other hand, the transposition is completely effective in the case of **far-end** crosstalk because the slightly higher currents induced on the energized side of the transposition are attenuated more in reaching the far end of the circuit than are the currents induced on the far side of the transposition.

The second reason why a single transposition is not effective in reducing crosstalk to the desired minimum is the phase change of the transmitted currents. In a long circuit, several wavelengths may be included in the propagation of a voice
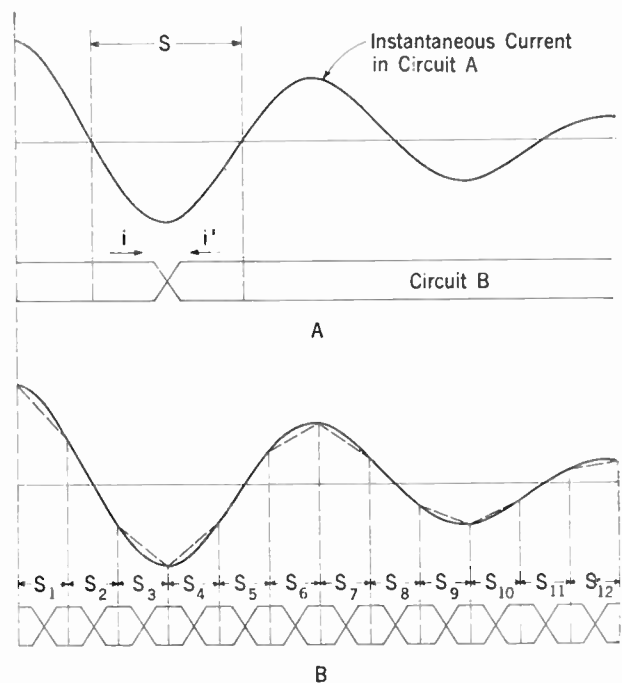
FIGURE 32-8

current from one end to the other. Since crosstalk is an induced effect, its instantaneous value in any small section S depends upon the position of S with respect to the cycle of current in the disturbing circuit. If S is so located that the current or voltage in it has a maximum value, either positive or negative, we cannot expect the crosstalk induced here to be neutralized by the crosstalk in some other similar section, which is located at a point in the line where the voltage or current has a value nearly zero at the same instant.

It is necessary, accordingly, that transpositions be installed at frequent intervals with respect to the wavelength of the propagated current. This is illustrated by Figure 32-8. Here if we assume the instantaneous current condition shown at A, it may be seen that in the section S the voltages induced in circuit B on the two sides of the transposition would be out of phase with each other by about 90°. The transposition would, therefore, not decrease the induced crosstalk. However, a number of transpositions within a single wavelength, as illustrated at B, will reduce the crosstalk to practical limits, although the crosstalk could be completely eliminated only by the use of an infinite number of transpositions. The dotted curve would then become identical with the solid curve.

For voice-frequency transmission, where the frequencies are relatively low and the wavelengths correspondingly long, it is not difficult to obtain a sufficiently good approximation to this condition. Where high-frequency carrier systems are used, on the other hand, the wavelengths are so short as to require closely spaced transpositions. In open wire lines, spacings as close as every second pole are used in practice where the wires are carrying frequencies up to 140 kc. The degree of effectiveness of such closely spaced transpositions is of course dependent upon accurate spacing of the poles themselves, and is finally limited in practice by economic rather than theoretical factors.

There is another fundamental consideration in connection with crosstalk that is of the first order of importance. In what has been said above regarding transpositions, it has been tacitly assumed that the four wires which we were considering were of the same gage and material—and particularly, that the two wires of each pair were electrically identical. The latter is of course ordinarily the case in practice. However, there is always a possibility in open wire lines of the series

resistance, or the insulation resistance, of one wire of a pair being slightly different from that of its mate due to imperfect joints, defective insulators, etc. When this occurs, the assumption made in connection with Figures 32-6 and 32-7 that the currents flowing in the two wires of the disturbing pair, or the induced voltages in the two wires of the disturbed pair, were exactly equal and opposite is no longer true.

Thus, an unbalanced series resistance in wire 1 of Figure 32-6 would have the effect of reducing the current in wire 1 as compared to the current in wire 2 in an amount depending upon the value of the resistance. Under these circumstances, it is clear that the opposing voltages induced in wires 3 and 4 would no longer be exactly equal, and there would therefore be some crosstalk despite the transposition. A resistance or leakage unbalance in the 3–4 pair would cause a similar result. In this case, any resistance unbalance would cause one of the induced voltages to be smaller than the other, with the result that some crosstalk current would flow through the terminal. It is extremely important, accordingly, that the two wires of every talking pair be so constructed and maintained as always to have identical electrical characteristics.
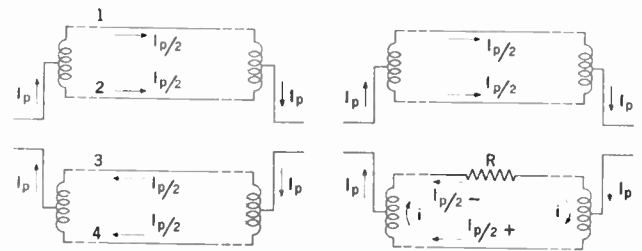
FIGURE 32-9

Resistance unbalance is particularly important when two pairs are used to create a phantom circuit. Here, under ideal conditions, exactly half of the phantom circuit current flows in each of the four wires, as shown in Figure 32-9. The phantom currents in wires 1 and 2 are then equal and in the same direction, and they therefore cause no current to flow through the terminal stations connected to these wires. The same is true of the side circuit made up of wires 3 and 4. A bad

joint, or resistance unbalance of any other kind, in any one of the four wires will reduce the current in that wire somewhat. As a result, the phantom currents in the two wires of the pair concerned will no longer be equal and an unbalance current will flow through the side circuit terminal. In other words, the phantom circuit will crosstalk into the side circuit, or vice versa. The effect of resistance unbalance in this situation is ordinarily much more serious than its effect on crosstalk between two side circuits or two non-phantomed circuits.

32.4 Crosstalk Reduction Practices

In considering practical methods for keeping the crosstalk in long toll circuits at a reasonable minimum, it is desirable first to consider the effects of certain basic design features of long circuits with respect to crosstalk. In general, these will apply equally to both open wire and cable facilities, and at either voice or carrier frequencies. One such important feature is the effect of the location of telephone repeaters on crosstalk. Thus, it is obvious that if two circuits are in close proximity at a point near a repeater station, and one circuit is carrying the high current levels coming from the output of a repeater while the other circuit is carrying the low current levels approaching the input of a repeater, the tendency of the first circuit to interfere with the second circuit is very great. The very small percentage of the current in the first circuit which may be induced into the second circuit will be amplified by the repeater on that circuit along with, and to the same degree as, the normal transmission. The best practical remedy for this condition, of course, is to avoid such situations by keeping circuits carrying high level energy away from low level circuits as much as possible. Where such physical separation between circuits is not feasible, differences in energy level between adjacent circuits can frequently be minimized by proper adjustment of repeater gains when the circuit is designed.

Another basic element of circuit design is that in most of the longer voice-frequency cable circuits and in all carrier circuits, the effect of near-end crosstalk is minimized by the use of separate paths for transmission in the two directions. In cable circuits, the wires carrying the transmission in the two directions are physically separated as much as possible by placing them in different

layers or segments of the cable; or, in the case of Type-K carrier circuits, in different cables. A comparable separation is obtained in open wire carrier circuits by using entirely different bands of frequencies for transmission in the two directions.

Furthermore, any near-end crosstalk occurring in spite of these physical separations is returned on the disturbed circuit to the output of an amplifier. Since the amplifier is a one-way device, the crosstalk can proceed no farther and does not reach the terminal of the circuit. Near-end crosstalk in such circuits is therefore of little importance, except in so far as it may be converted into far-end crosstalk by reflection from an impedance irregularity. To avoid this latter effect, it is essential that all circuit impedances be so matched as to eliminate important reflection possibilities.

Aside from the above techniques for avoiding crosstalk through circuit design methods, practical procedures differ considerably depending upon the type of facility. It is desirable, accordingly, to analyze the problem separately for open wire and cable facilities.

In the case of open wire lines, crosstalk reduction depends upon three principal factors—namely, wire configuration on the poles, transpositions, and resistance balance. Resistance balance is primarily a question of maintenance and ordinarily presents no great difficulty. The use of high-frequency carrier systems, with their much greater crosstalk possibilities, has led to the development of new configurations of open wire lines in which the wires of individual pairs are spaced closer together and the pairs are spaced farther apart. One standard pole-head configuration of this kind is illustrated in Figure 18-4, where it may be noted that the separation of the wires of each pair is 6 inches and the horizontal separation on the crossarm between any two wires of different pairs is at least 28 inches.

The basic principle of transpositions was outlined in the preceding Article. It was noted there that a large number of transpositions was needed in any long section of line to reduce crosstalk to the desired extent. In the entire discussion, moreover, only two pairs were considered. In practice an open wire line usually carries many more wires than this, and obviously there are crosstalk possibilities between any two pairs on such a line. These possibilities are greater between the pairs that are adjacent to each other, but all of the other possibilities are sufficiently large that they
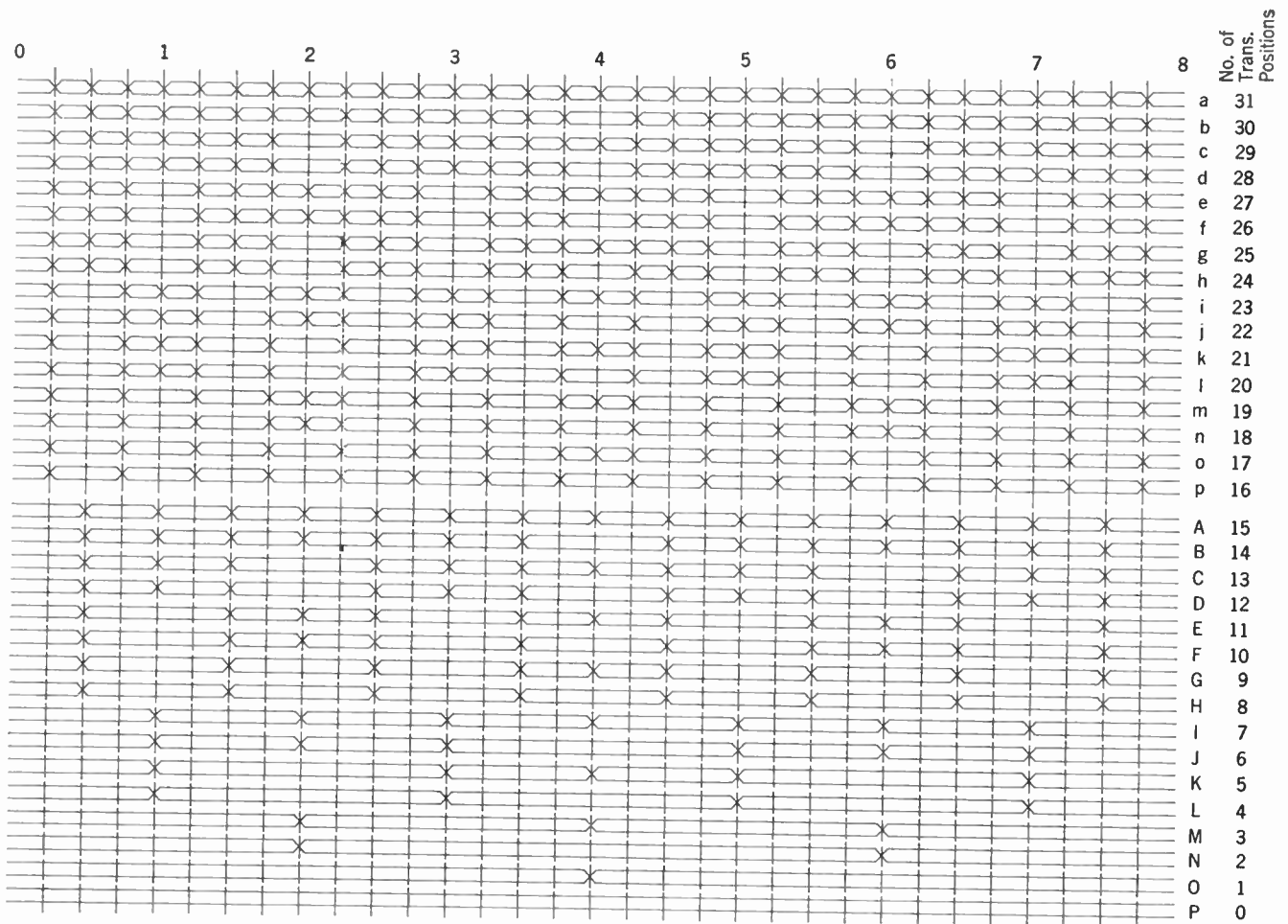
FIG. 32-10 STANDARD TRANSPOSITION CODE

must be taken into consideration in designing a transposition system for the line. A practical system must also guard against crosstalk between side and phantom circuits and between the phantoms themselves, when such circuits are used.

There is still another extremely important factor which has not been considered up to this time. This is the possibility of crosstalk from one circuit to another via a third circuit. In a line carrying many circuits, there are a large number of these tertiary circuits via which crosstalk might be carried from any one pair to any other pair. Even the hypothetical line that we considered in the first place, carrying only four wires, has two such tertiary circuits. These are the phantom circuit, made up of the two wires of one pair transmitting in one direction and the two wires of the other pair transmitting in the opposite direction; and the "ghost" circuit, made up of the

four wires acting as one side of a circuit, with a ground return. (Note that these circuits exist as tertiary crosstalk paths regardless of whether a working phantom circuit is actually applied to the four wires.) Needless to say, the presence of these tertiary circuits in a line complicates the problem of designing effective transposition systems. So much so, indeed, that no attempt can be made here to analyze this problem in detail.

Transposition systems for open wire lines are designed for unit lengths ranging from a few hundred feet to some eight miles. The purpose of the design is to approach as closely as possible to a complete crosstalk balance in each such unit section. Any number of sections can then be connected in tandem. The non-uniformity in the length of sections is the result of discontinuities in the line, such as junctions with other lines, wires dropped off or added, etc. It is naturally

[340]

desirable that such points of discontinuity should coincide with junctions between transposition sections, where the crosstalk is balanced out.

Figure 32-10 illustrates the fundamental transposition designs used in a section with 32 or less transposition poles. These fundamental types are frequently extended to include 64, 128, or even as many as 256 transposition poles per section, on lines to which high-frequency carrier systems are applied.

Physically, there are two standard methods for effecting transpositions between wires on pole lines. These, known as "point type" and "drop bracket" transpositions, are shown respectively in Figures 32-11 and 32-12. The former (point type) is widely used on lines carrying carrier systems because it does not change the configuration of the wires in the adjacent spans, as does the drop bracket type. Where very high frequencies are used, as in the Type-J carrier system, this becomes extremely important. In fact, the sensitivity of these carrier systems to crosstalk is so great that every possible effort has to be made to avoid even slight deviations in the amount of sag of the wires in the spans between poles.

Turning now to cable, the most striking feature of this type of facility with respect to crosstalk is that the conductors are crowded closely together. This is particularly true of the two wires of each circuit pair, which are separated by only thin coatings of paper insulation. As we have already seen, this close spacing of the two wires of a pair

in which equal and opposite currents are flowing tends to minimize the external effect of the electromagnetic field of the pair. Moreover, in the process of manufacture, the cable conductors are very thoroughly transposed by twisting the two wires of each pair together, by twisting the two pairs of each group of four wires together to form quads, and by spiralling the quads in opposite directions about the cable core. Cables are also so manufactured and installed that their conductors are practically free from series resistance unbalances or insulation leakages. On the other hand, the close spacing of many circuits within the cable sheath, as well as their proximity to the sheath itself, offsets the above advantages to a considerable extent.

FIGURE 32-13

At voice frequencies, magnetic induction (inductive coupling) between circuits in a cable is normally so small as to be of relatively little importance in causing crosstalk. The same cannot be said of electric induction (capacitive coupling). Despite the most careful manufacturing methods, the capacitance unbalances between cable conductors usually remain large enough to cause objectionable crosstalk in long circuits. This crosstalk is guarded against in practice by the use of additional balancing techniques when a toll cable is installed.

Voice-frequency crosstalk between circuits in different quads of a cable can be reduced to a satisfactory minimum at that time by splicing the successive lengths of cable in a more or less ran-

FIG. 32-11 POINT-TYPE TRANSPOSITION

dom manner such that no two quads are adjacent for more than a small part of their total length. This technique of course has no effect upon the crosstalk between circuits in the same quad. To reduce this crosstalk, it is necessary to measure the capacitance unbalances of each quad at the time of installation and then to correct such unbalances as are found large enough to be likely to cause serious crosstalk.

There are two principal methods of effecting this latter correction. One depends upon measuring the unbalances at several equally spaced splicing points within each loading section, and then splicing the quads together in such a way that a given unbalance in one section is counteracted by an equal and opposite unbalance in the adjacent section. This will perhaps be made clearer by referring to Figure 32-13 where the four wires of a quad are shown in cross-section, with the capacitances between the wires indicated by small capacitors. The wires marked 1 and 2 form one pair of the quad, and the wires marked 3 and 4 the other. (The capacitances between the pairs themselves are not shown because they have no effect on crosstalk.) The ideal condition in such a quad is that the values of all four capacitances A, B, C, and D shall be equal, and that capacitance E shall equal capacitance F and capacitance G shall equal capacitance H. In this case there is no unbalance within the quad and no crosstalk. However, if it is found, for example, that capacitance A in a certain quad of one section of the cable is too low, this quad can be spliced to a quad in an adjacent section of equal length in which capacitance A is too high by an approximately equal amount. The net unbalance of the connected quad over the two sections will thus be made to approach zero.

In the second method, the unbalances are counteracted in part by connecting small balancing capacitors into the circuits at one or two points in each loading section. This, combined with a limited number of "test splices" as above, effects the net result desired with greater accuracy, and reduces the number of capacitance unbalance tests that have to be made when a cable is installed. These balancing capacitors consist of short lengths of two parallel insulated fine-gage wires wound helically around a non-conducting core. Two terminals of this tiny capacitor are connected across the two line conductors whose capacitance it is desired to increase, and the other ends of the wires can be cut off at whatever point

is necessary to give the capacitor the precise value required. A large number of these capacitors can be included within the sleeve at a splicing point. In certain cases where the cable conductors are to be used for 4-wire circuits, it is practicable to balance the capacitances for a whole repeater section by adding capacitors of this type at one end.

The capacitance balancing methods outlined above have been found adequate in practice for keeping crosstalk to a tolerable minimum in voice-frequency cable circuits. When carrier systems are applied to cable circuits, the crosstalk problem becomes more severe. In this case, while capacitive coupling is still of consequence, inductive coupling becomes much more important as a cause of crosstalk. In fact, at the highest frequencies of the Type-K carrier systems, it predominates over capacitive coupling as a cause of crosstalk in the ratio of about 3 to 1. Accordingly, additional crosstalk reduction measures must be applied to cable conductors used for such carrier systems.

The crosstalk possibilities at these high frequencies are so great, in fact, that a number of basic changes in circuit design are required. In the first place, the carrier pairs are used for carrier transmission only. Next, the transmitting paths in the two directions are kept entirely separated by using separate cables for transmission East to West and West to East. The circuits in the two directions are likewise kept separated within the terminal offices and repeater stations, and shielded office wiring is used in all cases. This means that the energy levels of the carrier currents are approximately the same in all physically adjacent conductors, and that near-end crosstalk possibilities are completely eliminated (assuming that reflection effects have been properly guarded against).

Far-end crosstalk between carrier pairs is minimized by balancing out the capacitive and inductive couplings. In addition, special precautions are taken to prevent interaction crosstalk between carrier pairs via the voice-frequency pairs in the cable. Most effective in accomplishing this latter, is the complete transposition of the entire group of carrier pairs between the two cables at each repeater station. As may be seen from Figure 32-14, this automatically eliminates crosstalk via the voice-frequency pairs from the outputs of the amplifiers in the carrier pairs to the inputs of amplifiers in other carrier pairs. Carrier filters or noise suppression coils are also inserted in the
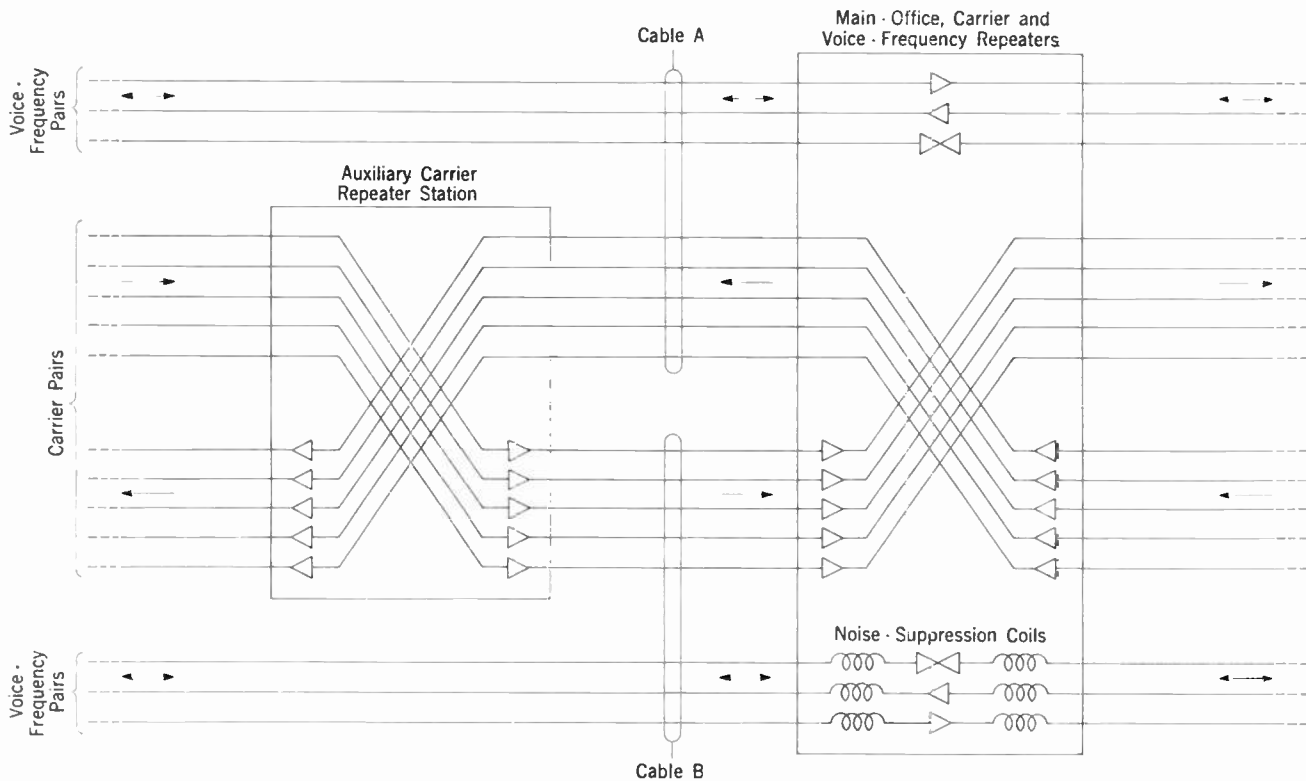
FIG. 32-14 ARRANGEMENTS FOR REDUCING NOISE AND CROSSTALK IN TYPE-K CARRIER SYSTEMS

voice-frequency pairs at voice-frequency repeater stations, and certain other points, to discourage the transmission of induced carrier frequency currents over the voice-frequency conductors.

The methods of balancing out capacitive coupling between the carrier pairs themselves are essentially the same as were discussed above in connection with voice-frequency transmission. In balancing out crosstalk due to inductive coupling, different methods must be used. The fundamental problem is to balance every carrier pair against every other carrier pair in the same cable, in each repeater section. The method used depends in effect upon counteracting the crosstalk currents with equal currents flowing in the opposite direction. Thus, if in a given disturbed circuit a crosstalk current is flowing in a clockwise direction, it is desired to set up an equal current in the circuit flowing in a counterclockwise direction.

This result can be effected by means of tiny transformers connected between each carrier pair and every other carrier pair. However, since it is necessary to control the magnitude of the artificially induced currents and also to cause them

to flow in either direction, depending upon the direction of the crosstalk current, the transformers must be designed so that the coupling between circuits can be adjusted and so that they can be poled in either direction. The method used to obtain this result is indicated schematically in Figure 32-15. Here, it may be noted that there are really two separate transformers, one having a reversed winding in the disturbing circuit so that a current, I, flowing in the disturbing circuit will induce oppositely poled voltages in the disturbed

FIG. 32-15 PRINCIPLE OF CROSSTALK BALANCING COIL

[343]

```
19-20  18-20  16-18  14-16  12-14  10-12   8-10   6-8    4-6    2-4    1-2
   18-19  16-20  14-18  12-16  10-14   8-12   6-10   4-8    2-6    1-4
 17-8   16-19  14-20  12-18  10-16   8-14   6-12   4-10   2-8    1-6      3-4
   16-17  14-19  12-20  10-18   8-16   6-14   4-12   2-10   1-8    3-6
 15-16  14-7   12-19  10-20   8-18   6-16   4-14   2-12   1-10   3-8      5-6
   14-15  12-17  10-19   8-20   6-18   4-16   2-14   1-12   3-10   5-8
 13-14  12-15  10-17   8-19   6-20   4-18   2-16   1-14   3-12   5-10     7-8
   12-13  10-15   8-17   6-19   4-20   2-18   1-16   3-14   5-12   7-10
 11-12  10-13   8-15   6-17   4-19   2-20   1-18   3-16   5-14   7-12     9-10
   10-11   8-13   6-15   4-17   2-19   1-20   3-18   5-16   7-14   9-12
  9-10   8-11   6-13   4-15   2-17   1-19   3-20   5-18   7-16   9-14    11-12
    8-9    6-11   4-13   2-15   1-17   3-19   5-20   7-18   9-16  11-14
  7-8    6-9    4-11   2-13   1-15   3-17   5-19   7-20   9-18  11-16    13-14
    6-7    4-9    2-11   1-13   3-15   5-17   7-19   9-20  11-18  13-16
  5-6    4-7    2-9    1-11   3-13   5-15   7-17   9-19  11-20  13-18    15-16
    4-5    2-7    1-9    3-11   5-13   7-15   9-17  11-19  13-20  15-18
  3-4    2-5    1-7    3-9    5-11   7-13   9-15  11-17  13-19  15-20    17-18
    2-3    1-5    3-7    5-9    7-11   9-13  11-15  13-17  15-19  17-20
  1-2    1-3    3-5    5-7    7-9    9-11  11-13  13-15  15-17  17-19    19-20
```

FIG. 32-16 METHOD OF CONNECTING CROSSTALK BALANCING COILS

tion becomes rather large since one coil is required for every possible combination of pairs. In practice, also, an additional coil is used for each quad to provide sufficient margin for balancing out side-to-side crosstalk. Thus, 20 pairs require a total of 200 coils, 40 pairs require 800 coils, and the maximum of 100 pairs requires 5000 coils. The coils are installed in unit panels arranged for balancing 20 pairs, and additional intergroup panels are added as successive 20 pair carrier groups are put into service. A special crisscross wiring arrangement, such as is indicated in Figure 32-16, is employed. This is necessary in order that the currents in any two pairs shall flow through the same number of coils before reaching the coil that balances these two pairs, thus insuring that the phase shift up to the balancing coil will be approximately the same on both pairs.

32.5 Noise

Voltages within the voice-frequency range, induced in a telephone circuit by electric power circuits, are manifested to a listener on the tele-

circuit. If the cores of the two transformers are centered as shown in the drawing, the induced voltages will be exactly equal and the net effect on the disturbed circuit will be nil. By moving the two cores as a unit in either direction, however, one or the other of the induced voltages can be made to predominate. Thus, if the cores are moved to the left, voltage e_1 will be increased while voltage e_2 will be decreased by a like amount. The result will be a current flowing in a counterclockwise direction in the disturbed circuit. On the other hand, moving the cores to the right will cause a clockwise current in the disturbed circuit, the value of which will depend upon the extent of the movement of the cores.

In practice, the balancing coils are designed to have a mutual inductance ranging from approximately $+1.6$ to -1.6 microhenrys for the two limiting positions of the cores. The coils are mounted in cylindrical containers arranged for rack mounting. The position of the coil cores is controlled by a screw by means of which the core can be moved through its maximum travel of $\frac{1}{2}$ inch in about 16 complete turns.

In using these coils to balance out crosstalk, measurements of the inductive coupling between each pair of conductors must be made and each coil adjusted to counteract this coupling. In a cable containing a large number of carrier pairs, the number of coils required at each repeater sta-

CABLE BALANCING CAPACITANCES—OLDER TYPE SHIELDED TWISTED PAIR AT LEFT—NEWER TYPE SPIRALLY WOUND PAIR AT RIGHT—UNIT USED AT END OF REPEATER SECTION IN CENTER
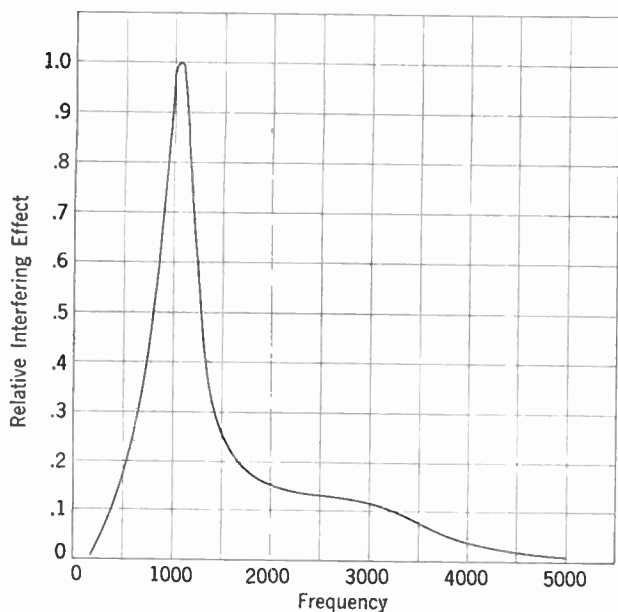
FIG. 32-17 RELATIVE INTERFERING EFFECT OF NOISE AT
DIFFERENT FREQUENCIES

phone circuit as noise. In many cases, crosstalk currents may also appear merely as noise. This is particularly true in the case of cable circuits where any crosstalk heard is likely to come simultaneously from a considerable number of other circuits, and appears to the listener on the disturbed circuit as a special form of noise, called "babble". In other words, it is just an unintelligible conglomeration of speech sounds coming from a large number of sources.

The disturbing effect of noise to a listener depends first, of course, upon its volume. It also depends upon the frequency of the noise currents. Figure 32-17 shows the results of tests that have been made to determine the relative disturbing effects of various noise frequencies. It will be noted that the disturbing effect peaks up rather sharply in the neighborhood of 1100 cycles. Where noise is of appreciable volume—particularly in the more sensitive frequency range—it it naturally annoying to the telephone user and may seriously reduce the intelligibility of conversation. It is accordingly necessary to keep the noise in working telephone circuits below those limits where its interfering effect on conversation will be important.

Since noise is essentially an induced effect like crosstalk, similar measures are used to counteract it. Careful resistance balancing of the telephone

conductors, the use of transposition systems, and other measures taken to avoid crosstalk, are likewise effective in reducing noise. However, such measures alone may be inadequate to keep noise within the desired limits. This is a result of the fact that paralleling power lines are the principal source of noise, and the power carried over such lines is greater by tremendous percentages than that carried over any telephone circuit.

Of course, the usual fundamental frequency of power transmission is 60 cycles and this frequency is too low to interfere appreciably with telephone transmission. Unfortunately, however, the currents transmitted over the power line usually include several harmonics of the fundamental frequency, and these may lie well within the range of telephone frequencies. In high tension power lines, such harmonics may have energy values as high as 10,000 watts and when this is compared with the energy in the telephone circuit, which may be as low as .00001 watt, it will be evident that, even for a considerable physical separation between the power and telephone conductors, the danger of serious interference is great.

The ideal way of eliminating such interference is to avoid any parallels of small separation and appreciable length between power and telephone lines. This is done whenever practicable. In many cases, however, parallels with fairly close separation, such as lines on the opposite sides of a highway, cannot be avoided. In these cases, it is frequently necessary to make use of certain measures additional to balancing and transposing the telephone conductors. These usually require the cooperation of the power companies. The particular techniques to be used vary somewhat with each situation but include such measures as rearrange-
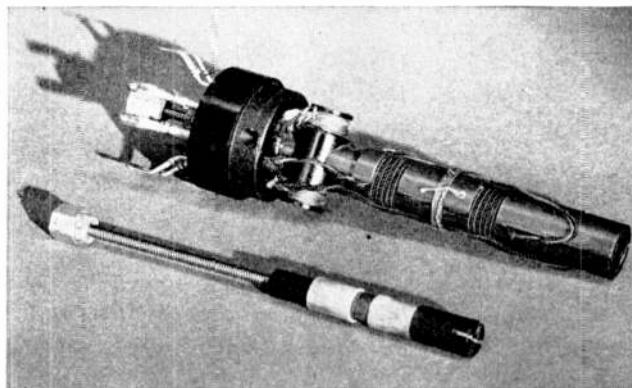
CONSTRUCTION OF CROSSTALK BALANCING COIL

[345]
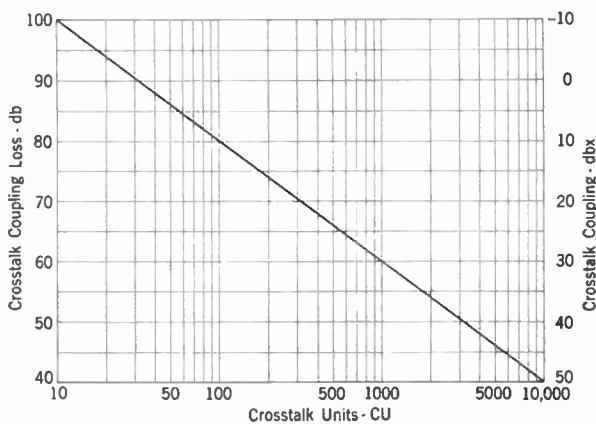
FIG. 32-18 RELATIONS BETWEEN CROSSTALK
MEASURING UNITS

ments of the transformer connections in the power circuits, or the insertion of filters to reduce harmonics. Other methods frequently employed include changes in the configuration of the power wires on their poles to make for better electrical balance, and transpositions of the power wires. Such power line transpositions have essentially the same effect in balancing out the magnetic fields as do transpositions in a disturbing telephone circuit.

32.6 Crosstalk and Noise Measuring Units

The ideal objective of the various methods discussed above for counteracting crosstalk and noise induction in telephone circuits is, of course, to eliminate their effects altogether. In practice this ideal is rarely attained. But certain practical limits are established, and every reasonable effort is made to keep the crosstalk and noise below these limits. In designing and maintaining circuits, therefore, it is desirable to be able to make definite quantitative measurements of both crosstalk and noise. As in any other kind of measurement, this requires the establishment of definite units.

The measure of either crosstalk or noise that would be of major significance as a practical matter is the extent of the interference or annoyance to which a listener on a disturbed circuit is subjected. Since such a measure is obviously affected by numerous subjective factors, it is clear that completely objective quantitative measurements of crosstalk and noise effects are practically impossible. It is possible, however, to make precise quantitative measurements of the **crosstalk coup-**

ling between a given sending point on a disturbing circuit and a given receiving point on a disturbed circuit. Essentially this is simply the measurement of the transmission loss between the two points, and like any other transmission measurement it may be made at one or more frequencies as desired. Such a measurement gives a value of what is known as "crosstalk coupling loss" in db. A more commonly used measure of crosstalk coupling employs a unit designated dbx, which expresses the coupling in db above "reference coupling". Reference coupling is equivalent, broadly speaking, to a crosstalk coupling loss of 90 db and is formally defined as "the coupling which would be required to give a reading of zero dba on a 2-type noise measuring set connected to the disturbed circuit when a test tone of 90 dba (using the same weighting as that used on the disturbed circuit) is impressed on the disturbing circuit". Another unit is sometimes used for measuring crosstalk coupling. This is the "crosstalk unit" abbreviated CU. The number of crosstalk units representing any given coupling is 10^6 times the ratio of the current or voltage in the disturbed circuit to the current or voltage in the disturbing circuit at the two points under consideration; or, if the circuit impedances are not the same, 10^6 times the square root of the power ratio. The relationships between the three measures of crosstalk coupling are shown graphically in Figure 32-18.

For measuring noise, a basic reference point has been selected, which is equal to 10^{-12} watts of 1000-cycle power. This corresponds to 90 db below 1 milliwatt (-90 dbm). Noise may then be measured in terms of number of decibels above this reference point. However, the interfering effect of noise on a listener varies with both the power level and the frequency; and the relative importance of the components of noise at the different frequencies must be taken into consideration in determining the total amount of interference. The interfering effect also varies according to the sensitivity of the receiving device that converts the noise currents into audible sound. For these reasons, in measuring noise, it is desirable to employ "weighting networks" which act to integrate the noise power over the voice-frequency range by giving each small band of frequencies a weighting proportional to its contribution to the total interfering effect. Different weighting networks may be used with different receiving devices. Even so, equal values of db reading will not

necessarily indicate equal interfering effects without some adjustment of the calibration constants. In practice, an adjusted unit designated dba is employed, which measures the acoustic interfering effect of the frequency-weighted noise energy. Equal values of dba measured across any receiving device, with proper weighting used, should indicate approximately equal interfering effects.

CHAPTER 33

ALTERNATING-CURRENT MEASUREMENTS

33.1 Alternating-Current Meters

Meters of the galvanometer type, in which the current to be measured flows through the windings of a movable coil suspended in the field of a permanent magnet, are commonly used for the measurement of direct currents. Such meters are not suitable for measuring alternating currents because they would tend to indicate the instantaneous values of the rapidly changing current, both positive and negative.

For the measurement of alternating currents of relatively low frequency (up to a few hundred cycles), direct-reading meters employing the **dynamometer** principle are frequently used. These are similar in external appearance to the usual d-c meters, but the permanent magnet of the galvanometer type meter is replaced by a stationary coil through whose windings the current to be measured flows in series with the windings of the movable coil. Current reversals thus occur simultaneously in both coils and the interaction between their magnetic fields is such as to cause the movable coil to rotate in only one direction regardless of the direction of the current in the windings. The deflection of the movable coil is proportional to the **effective** value of the current through its winding. Within the frequency limits indicated, this type of instrument may be used to measure alternating current, voltage or power provided that the values are of sufficient magnitude to register. For power measurement, however, one of the coils must be connected in series so that the current in it varies as the line current, while the other coil is connected across the circuit so that the current in it is proportional to the voltage. This automatically takes care of any phase difference between current and voltage, permitting the meter to register in watts.

The direct-reading a-c meter is ordinarily not adequate for measuring current, voltage and power at the higher frequencies generally employed in communications work because of the extremely small values of the quantities involved and other difficulties inherent in the design of a-c meters at high frequencies. Since it is possible to build d-c meters with a much higher degree of

sensitivity, the great majority of a-c measurements in communications work are made indirectly by rectifying the alternating current and then measuring the rectified current with a d-c meter. If the a-c quantities are still too small to register on a meter, an appropriate amplifier may be inserted in front of the rectifying device. Where it is not necessary to use an amplifier, the conversion of small alternating currents to proportional direct currents is usually effected by a thermocouple. The arrangement is illustrated schematically in Figure 33-1 where the alternating current to be measured, flowing through the wire ab, heats the junction H of two dissimilar metal conductors. Due to thermo-electric effect, this produces a difference of potential between the two dissimilar conductors, which causes a direct current to flow through the d-c meter. When the meter is properly calibrated by means of the potentiometer P, this current is exactly equal to the effective value of the applied alternating current. Where the a-c quantities are so small as to require the use of an amplifier for their measurement, rectification is usually effected by means of either disc or point-contact rectifiers.

For most purposes, a-c measurements in communications work are not made in terms of volts or amperes but in logarithmic units such as the db or dbm. Many of the meters used are accordingly designed with scales reading in these units. The ordinary d-c ammeter is designed to have equal spacing on its scale for equal increments of current. Since the movable coil rotates in an air-gap of uniform flux density, the torque on the coil is directly proportional to the current. If such a meter is calibrated to read in decibels, the points on the scale for successive units above or below
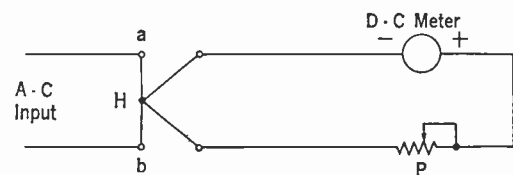
FIG. 33-1 THERMOCOUPLE A-C METER

[348]
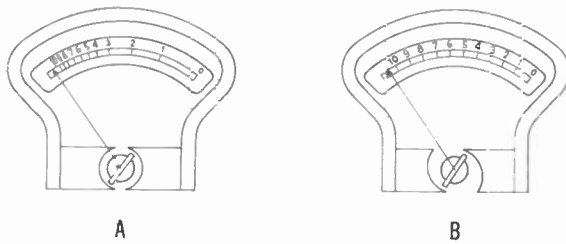
FIG. 33-2 DECIBEL METER

zero will be spaced increasingly close together, as shown in Figure 33-2(A), because the decibel is a logarithmic unit. To obtain the even spacing on the db or dbm scale that is generally desirable, an ammeter may be made as shown in Figure 33-2 (B), with the pole-pieces of the permanent magnet cut in an eccentric pattern so that the flux across the air-gap increases logarithmically as the needle moves away from the zero position. The advantages of such a design are somewhat offset by the fact that this meter requires a higher operating current because of the lower flux at the zero position.

It should be noted that the zero of a db or dbm scale does not have the same significance as that of an ordinary linear scale because the decibel by definition is the logarithm of a ratio, and zero is the logarithm of 1. When an ordinary ammeter reads zero, it means that no current is flowing in its winding. When a db meter reads zero, it means only that the ratio between the power being measured and some reference power is 1. This reference power, as we know, is 1 milliwatt for measurements in dbm. Under other conditions, the relationship between meter reading and reference power may be complicated.

Although not explicitly stated, it has been assumed in the foregoing that the alternating currents and voltages dealt with were in the form of sine waves. When this is not the case, measurements in terms of the units discussed above may have little meaning. Speech waves, for example, vary with time in a complex manner and it is not possible to measure their precise values in terms of watts or decibels except on an instantaneous basis. It is frequently desirable, however, to know the overall average strength of transmitted speech or music waves. A device known as a **volume indicator** may be used for this purpose. This is essentially a high impedance voltmeter made up of an attenuator, a copper-oxide rectifier, and a d-c milliammeter having specified dynamic characteristics, which may be bridged across a line without appreciable effect on normal transmission. The meter is calibrated in terms of VU. This is a logarithmic unit that measures strength or volume above or below a specified reference level. In general, this reference level, designated 0 VU, indicates no precise electrical quantity, but the volume indicator is calibrated to read 0 VU on 1 milliwatt of 1000-cycle power dissipated in a 600-ohm resistance. The VU represents the same power ratio as the db and the volume indicator may therefore be used to measure transmission losses or gains when the current being measured is a 1000-cycle sine wave, although its primary purpose is to measure the volume of complex waves.

Another example of measurements where the ordinary electrical units cannot be practicably applied is presented by the video signal of television program transmission. This signal is not only non-sinusoidal, but its positive and negative peaks are usually unequally displaced from the zero axis. Under these conditions, a measurement of the signal amplitude in ordinary power terms is meaningless. As has been indicated earlier in this book, video signal amplitude is commonly defined in terms of peak-to-peak voltage—that is, the voltage difference between the maximum swing of the signal in one direction and that in the other direction. For general measurements of video signal strength, it is convenient to use a unit known as the dbv. This expresses the ratio in db of the peak-to-peak voltage at any point to a reference voltage of 1 volt peak-to-peak. The 1 volt peak-to-peak reference voltage is zero dbv and for a sine wave its value is equivalent to an effective value of 0.354 volts.

33.2 The Cathode-Ray Oscilloscope

The cathode-ray oscilloscope is fundamentally a voltage measuring device that is particularly useful for observing and measuring a-c voltages of non-sinusoidal waveshape. Its basic element is an electrostatic cathode-ray tube, which, because of the almost negligible inertia of its electron beam, makes possible the visual observation of the characteristics of extremely rapid voltage changes, provided that they recur with sufficient frequency to produce a persistent light pattern on the screen of the tube.

[349]
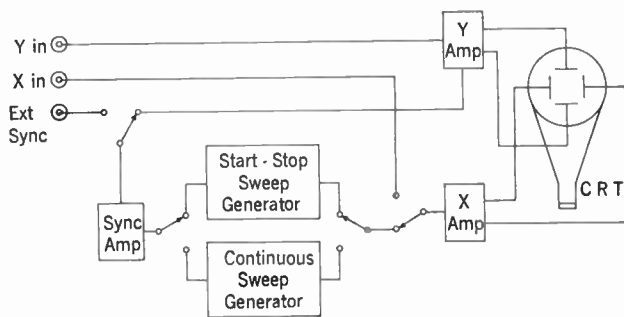
FIG. 33-3 BLOCK DIAGRAM OF CATHODE-RAY OSCILLOSCOPE

Figure 33-3 shows in block diagram the essential circuit elements that are ordinarily associated with the cathode-ray tube to make up an oscilloscope that is adaptable for the observation and measurement of signals of many types. For most observations, the voltage under test is applied to the vertical or Y deflecting plates of the tube while a linear sweep voltage of the same frequency as the voltage under observation is applied to the horizontal or X deflecting plates. The amplifiers designated Y and X are used to bring the deflecting plate voltages up to the values necessary to cause the desired swings of the electron beam. There are two sweep generators, both usually of the multivibrator type. One is free-running to give a continuous saw-tooth wave output; the other is of the start-stop type, which stops at the end of each saw-tooth cycle until started by a "triggering" signal. The latter makes it possible to obtain a steady trace of a signal whose recurrence is not uniform with time. Switches are provided as shown in the diagram for connecting either generator into the circuit. Both sweep generators are arranged with appropriate controls for adjusting their frequencies over a wide range.

In order to hold the sweep circuits in the exact synchronism that is necessary to produce a stationary pattern on the tube screen, synchronizing pulses are applied to the multivibrators through a SYNC amplifier, as indicated. These pulses are ordinarily obtained internally from the output of the Y amplifier, but provision is made for picking them up from an external source by operating the switch to the EXT SYNC position. When a linear time scale is applied to the X plates of the tube, as discussed above, the screen gives a direct picture of the shape of the wave under investigation. By placing an appropriately calibrated grid scale on the face of the tube, reasonably accurate quantitative measurements may also be read directly from the picture.

For most of the purposes for which it is used in communications work, the oscilloscope is operated with a time scale applied to the X axis. It is possible, however, to apply an independent external voltage to the X plates as well as to the Y plates. Such a procedure is employed when the oscilloscope is used for measuring frequencies, as discussed briefly in the following Article.

33.3 Frequency Measurements

Methods for measuring the frequency of sine wave alternating currents depend upon the frequency range under investigation and the degree of accuracy required. At frequencies in the "power" range, up to about 200 cycles, direct-reading meters employing the principle of mechanical resonance are commonly used. Such meters have a comb-like arrangement of metal reeds of differing lengths, each of which has a different natural period of mechanical resonance. The comb is mounted in the gap of an electromagnet to the windings of which the voltage to be measured is connected. A scale above the ends of the reeds is calibrated in cycles to show the frequency of the reed which vibrates with maximum intensity. At higher frequencies, the principle of electrical resonance is quite generally employed in frequency measurements. Any resonant circuit with adjustable capacitance or inductance values can be used for determining frequency. One illustration of such an application is the wave meter shown in Figure 33-4. In circuit A, if an inductance L and a capacitance C are so adjusted as to give resonance, the meter M will give a maximum reading. Similarly, circuit B will be resonant when maximum sound is heard in the telephone receivers which are connected in series with a rectifying device, D. To know the frequency at the resonant point, it is of course necessary that the variable capacitor be previously calibrated in such terms.
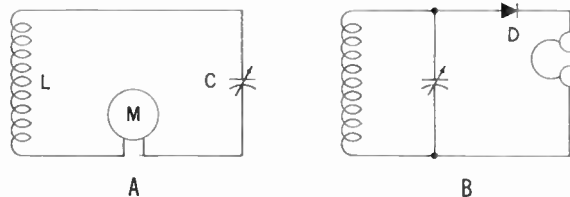
FIG. 33-4 WAVE METERS
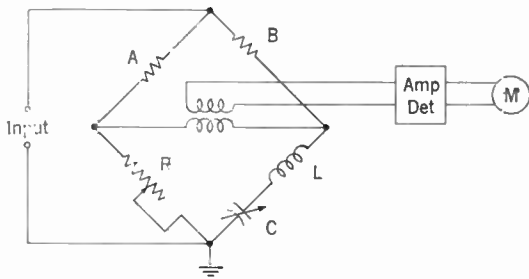
FIG. 33-5 BRIDGE-TYPE FREQUENCY METER

Wave meters of such comparatively simple design are widely used in connection with radio circuits where frequency determination of precise accuracy is rarely necessary. There are many situations in the voice and carrier frequency ranges where frequency measurements accurate to a few cycles, or even to one cycle, are desirable. One device employed for measuring frequencies in these ranges is the bridge circuit shown schematically in Figure 33-5. This also makes use of the principle of electrical resonance but employs a null method of measurement. The bridge arms A and B are of equal resistance and when the variable capacitor C and the variable resistor R are adjusted so that the meter registers no current, the LC arm of the bridge is in resonance with the input frequency. The scales of R and C are factory calibrated and a table is supplied with the bridge from which the frequency values corresponding to their settings may be read. The accuracy of this device of course depends upon the original accuracy, and the permanency, of this calibration.

Where greater precision is necessary, a meter employing a cathode-ray oscilloscope may be used. The essential elements of this device are shown in the block diagram of Figure 33-6. Its principle depends upon comparing the voltage under test with the output of a precisely calibrated variable oscillator by applying the two voltages respectively to the vertical and horizontal plates of a cathode-ray tube. When the two voltages have precisely the same frequency, a **stationary** pattern will appear on the screen of the tube. As was pointed out in Article 22.3, the form of the stationary pattern will depend upon the phase relationship of the two voltages, being a straight diagonal line when the voltages are in phase or 180° out of phase; a circle when the voltages are 90° out of phase; and elliptical at other phase angles. When such a stationary 1-to-1 frequency

pattern is obtained on the oscilloscope screen, the frequency adjusting dials of the variable oscillator will read the frequency directly to 1 cycle.

The stability of the bridge-type variable oscillator is very high. Nevertheless, it is necessary to calibrate it each time the meter is placed in operation and at intervals of about an hour during operation, if maximum precision is required. For this purpose, the meter includes a crystal standard whose known single frequency may be compared in the oscilloscope with several key frequencies in the variable oscillator's range, that are related to the fixed standard frequency by integral units. In these calibrating comparisons, the oscillator circuit constants are adjusted by "trimmer"
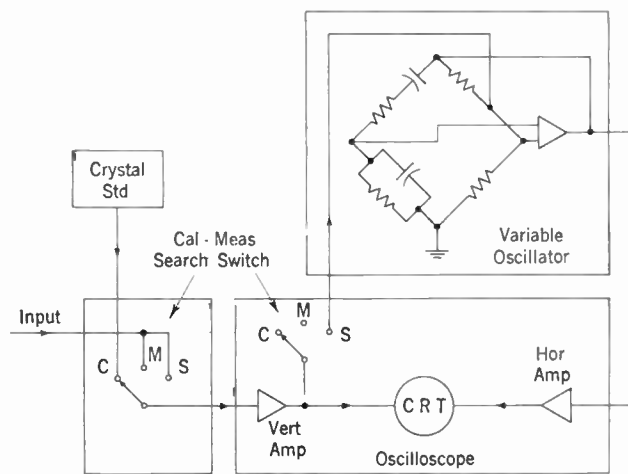
FIG. 33-6 OSCILLOSCOPE CIRCUIT FOR FREQUENCY MEASUREMENTS

resistors, if necessary, until a stationary pattern is obtained on the screen. Since there is only one crystal standard frequency, these stationary patterns will not represent 1-to-1 ratios but other integral relationships. Figure 33-7, for example, shows the patterns for a 10-to-1 and a 4-to-1 relationship, which correspond to oscillator frequencies of 400 and 1000 cycles respectively when the crystal standard is 4000 cycles.
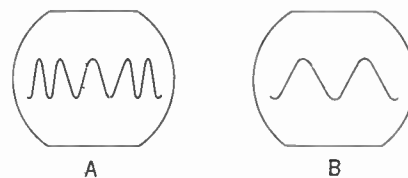
FIG. 33-7 OSCILLOSCOPE FREQUENCY PATTERNS
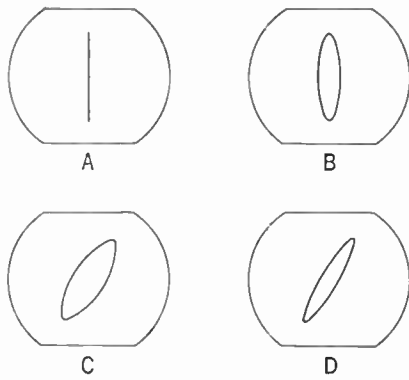
FIG. 33-8 SEARCH OSCILLOSCOPE PATTERNS

33.4 Impedance Measurements

Many times throughout this text, attention has been called to the necessity for matching impedances at junction points of communication circuits in order to eliminate unnecessary transmission losses or other undesirable effects. This makes it necessary, for practical maintenance purposes, to have available a device by means of which impedances can be accurately measured.

Figure 33-9 indicates the principle of a simple bridge circuit widely used in the telephone plant for measuring impedances in the voice-frequency range between 100 and 3000 cycles. As shown,

When the frequency of the voltage being measured is not even approximately known, the meter circuit can be arranged to search for the unknown frequency by operating the switch to the contacts marked S. Operation of the search switch converts the oscillator to a tuned amplifier. The amplifier output will then be negligible until and unless its tuning circuit is adjusted to a frequency that is close in value to the unknown input frequency. As long as the frequencies are far apart, only a straight vertical line will appear on the oscilloscope screen. When the tuning of the amplifier approaches the unknown frequency, this vertical line will take the form of an ellipse that will bend over and approach closer and closer to a straight diagonal line, as shown in Figure 33-8. At this point, the switch may be thrown back to M and the exact frequency determined.

FIG. 33-10 CIRCUIT OF A LINE IMPEDANCE BRIDGE

the unknown impedance is connected in one arm of the bridge and the balancing arm consists of a variable resistor and a variable inductor (inductometer) in series. Arms R_a and R_b are resistors of equal value. Measuring current is supplied from a variable oscillator capable of delivering satisfactory waveshape and output through the range of voice frequencies for which the bridge is designed. The values of R and L, when they are adjusted so that there is no current in the telephone receiver, will be equal to the corresponding values of the unknown impedance. The circuit as

FIG. 33-9 SIMPLE IMPEDANCE BRIDGE

[352]
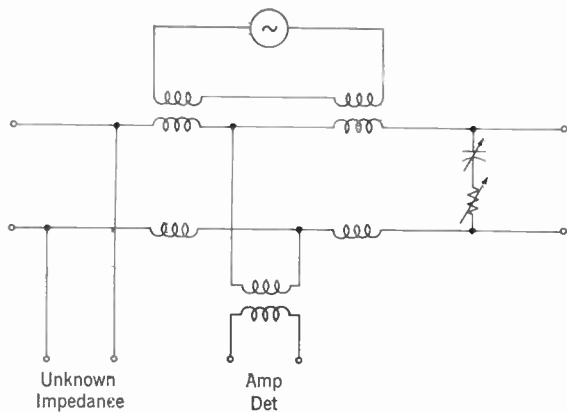
shown in the diagram could measure only an inductive impedance. The practical circuit, however, is arranged so that the variable inductor may be switched into the other arm of the bridge in series with the unknown impedance. When the bridge is balanced in this condition, the inductometer in effect gives a measure of negative inductance, which is equivalent to capacitance. The variable units are actually calibrated to read resistance in ohms and inductance in millihenries, but the readings may readily be converted into reactance and impedance values by the application of basic a-c equations.

A detailed wiring diagram of the bridge is given in Figure 33-10, which shows a number of features not indicated in Figure 33-9 but necessary to the practical operation of the bridge. Thus, the fact that it is not possible to build an inductometer that may be adjusted to zero inductance requires the use of a second fixed inductor which also may be switched from one arm of the bridge to the other. Zero inductance may then be secured by switching the fixed inductance to one arm of the bridge and the inductometer to the opposite arm so that the fixed inductor neutralizes its value on the scale of the inductometer. Since either inductor may be switched to either arm, the total reactance values that can be measured range from zero to $\pm 2\pi f \ (L_a - L_b)$, where L_a and L_b are the values of the two inductances. The diagram also shows two compensating resistors, whose values are equal to the resistances of the inductors. These are connected to the switch contacts in such a way as to be always in the opposite arm from the corresponding inductor, thus neutralizing the resistances of the inductors.

Other bridge designs, operating on a basically similar principle, are used for impedance measurements at higher frequencies. One of these, which is satisfactory for measurements between 1800 and 35,000 cycles, is shown schematically in Figure 33-11. The bridge here is the familiar hybrid coil. When the unknown impedance connected to the "line" side of the coil is matched by the adjustable impedance connected to the "net" side of the coil, voltage applied to the series winding from an oscillator will produce no current in the bridge connection to the amplifier-detector. It will be noted that the reactance adjustment in this circuit is made by means of a variable capacitor rather than an inductometer. If the reactance of the unknown impedance is inductive, the variable capacitor is transferred by an appropriate switch to the line side of the coil in series with the unknown impedance. Another bridge, designed for making measurements between 1 and 100 kc, is shown in Figure 33-12 in simplified schematic. This bridge differs from the usual circuit in that the ratio arms are four pairs of equal resistances, and the variable and unknown impedances are connected between mid-points of opposite pairs. The impedance is measured when the bridge is balanced in terms of resistance and capacitance in parallel rather than in series, and switches are provided to transfer the variable elements to the opposite side of the bridge if this should be necessary to secure balance.

One of the major uses of the impedance bridge in practical communications work is the location of impedance irregularities in long wire circuits. The impedance of a long line that is free from
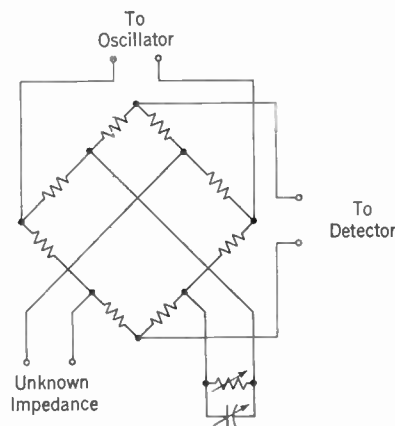
FIGURE 33-12

irregularities and terminated in its characteristic impedance, when measured over a wide band of frequencies, will appear as a smooth curve over the measured frequency range. If, however, there is an impedance irregularity along the line, such as might be caused by a defective or improperly located loading coil, some part of the energy applied to the line at the sending end will be reflected back from the point of irregularity. The reflected wave will add to or subtract from the initial applied wave, depending on its phase relationship when it reaches the sending end. The sending end impedance will be affected accordingly. The phase of the reflected wave with respect to the initial wave of course depends on the time it takes to travel from the irregularity to the sending end or, since the velocity of propagation is a constant for a particular type of facility, on the distance from the irregularity to the sending end.

The measured resistance and reactance of the line under these conditions, when plotted over a range of frequencies, will produce curves of the type illustrated in Figure 33-13. At points M, N, and O, the measured resistance values are maximum, indicating that the reflected current, when it reaches the sending end, is exactly opposite in phase to the initial current; while at P, Q, and R, the two currents are exactly in phase. Adjacent humps, either maximum or minimum, occur at frequency intervals corresponding to a change of one-half wavelength in the distance from the sending end to the point of irregularity. The distance may, therefore, be determined by the use of the

following equation:—

$$d = \frac{W}{2(f_2 - f_1)} \qquad (33{:}1)$$

where d is the distance to the irregularity, W is the velocity of propagation, and $f_2 - f_1$ is the average frequency interval between adjacent humps of the curve.

Where a large number of impedance-frequency measurements are to be made, particularly when the measurements need to be made quickly because the fault condition causing the irregularity is not constant, there are advantages in using a rectifying bridge of the type shown in Figure 33-14 that employs only one variable element. In this bridge, the unknown impedance is balanced by the variable resistance R and the bridge will therefore measure the magnitude but not the angle of the impedance. Current from the oscillator can flow
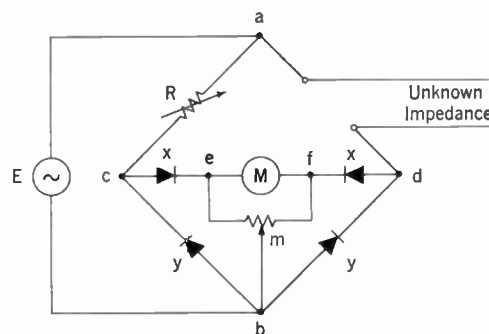
FIG. 33-14 RECTIFYING-TYPE IMPEDANCE BRIDGE

FIG. 33-13 IMPEDANCE CHARACTERISTIC OF CIRCUIT WITH SINGLE IRREGULARITY

in the meter circuit M only during the half-cycle when a is positive. In the negative half-cycle, current will flow through paths bca and bda but no current can flow in the meter circuit since any voltage applied across the rectifiers X in series with the meter will be of the wrong polarity for either one or the other of them. The bridge, therefore, acts as a half-wave rectifier. During the positive half of the cycle, current from E divides at a and flows through the paths $acemb$ and $adfmb$. If the resistances of these two paths are equal, there is no potential difference between e and f and no current will flow through the meter. The bridge is calibrated by adjusting the potentiometer m so that the meter will read zero when
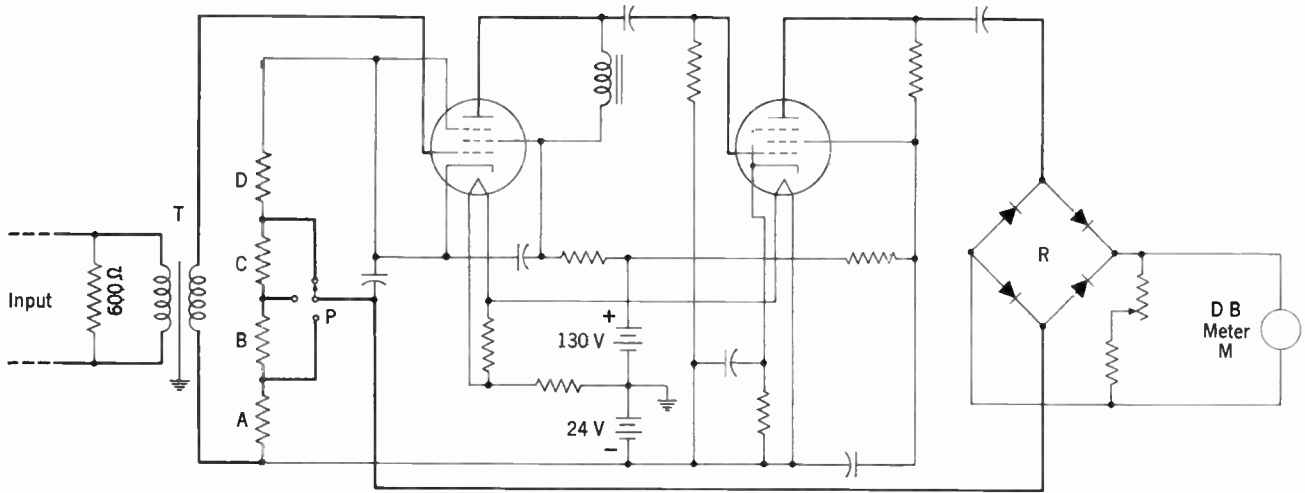
FIG. 33-15 DIRECT READING TRANSMISSION MEASURING SET WITH AMPLIFIER

equal resistances are inserted in the arms *ac* and *ad*. This corrects for any differences that may exist between the resistances of the two *X* varistors. The calibrating resistances should be of a value comparable, as nearly as possible, to the magnitude of the unknown impedance to be measured.

33.5 Transmission Measurements

Most widely used of the many types of measurements required in communications work are those known as transmission measurements. These are measurements of the ratio of the power at the receiving end of a transmission line to the power applied to the transmitting end. As is well known, they indicate the loss or gain of a circuit in terms of decibels or comparable logarithmic units. Two basic methods of making transmission measurements are commonly employed. The first is a direct method in which a known amount of power (generally 1 milliwatt) is applied to the sending end of the circuit under test and the power at the receiving end is measured by a direct-reading
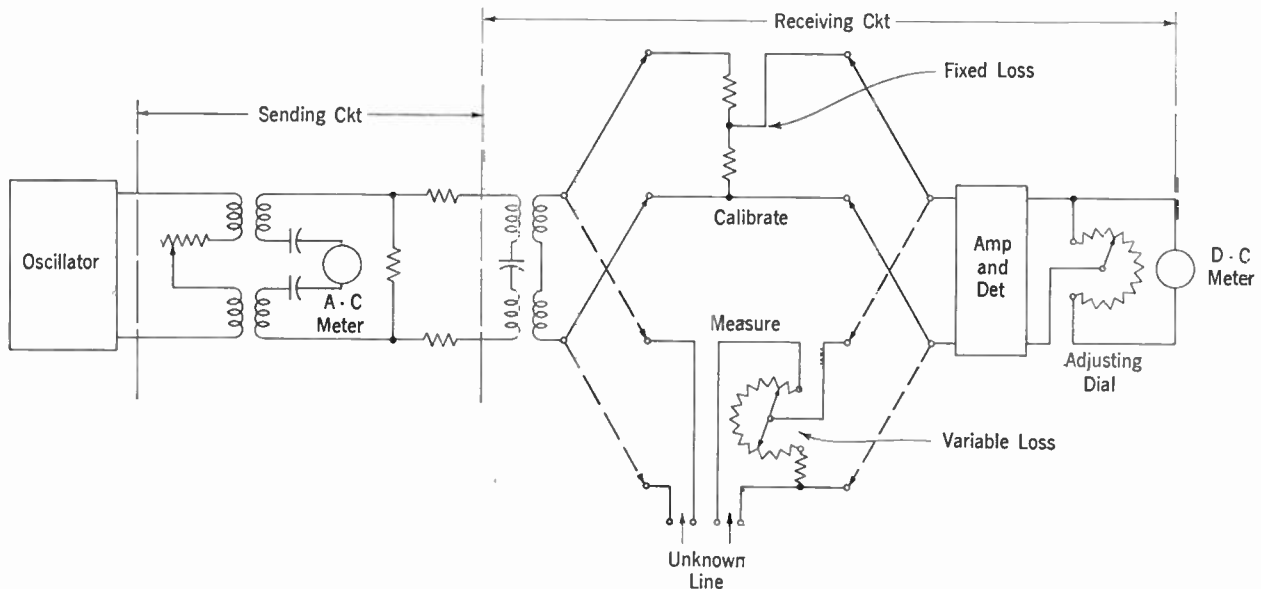
FIG. 33-16 PRINCIPLE OF TRANSMISSION MEASURING SET

meter in terms of db or dbm. This is obviously the simpler method and is used wherever practicable. In situations where it is not feasible to supply a known fixed power at the sending end of the circuit, a comparison method is used in which the loss or gain of the circuit under test is measured by comparing it with a known, calibrated loss or gain.

For routine checking of telephone circuits, transmission measurements are usually made at a single frequency of 1000 cycles and, in most cases, the direct method of measurement is employed. Fixed testing power of 1 milliwatt is supplied at the sending terminals from a 1000-cycle source of power, which consists of a small magneto-generator. At the receiving end, the power is amplified, rectified by copper-oxide varistors, and supplied to a d-c meter reading directly in db or dbm. The detailed circuit arrangement is shown in Figure 33-15. Where measurements at frequencies other than 1000 cycles are required, the same receiving circuit may be used but the sending power is furnished by an appropriate variable oscillator. To insure that the test power is at 1 milliwatt, the oscillator output must be calibrated against a fixed 1000-cycle generator output for each series of measurements at other frequencies.

In situations where a fixed known testing power source is not available, as would ordinarily be true for example in the case of portable transmission measuring sets, the comparison method mentioned above may be employed. The general principle of this type is illustrated in Figure 33-16. The set is first calibrated by connecting a voltage to a fixed artificial line which causes a definite known loss. The entering current, after passing through this line, is amplified and rectified and passed through a potentiometer to a d-c meter. The value of the applied voltage is then adjusted to such a value as to give any desired deflection of the d-c meter, usually mid-scale. After calibrating, connections are changed so that the same voltage is applied to a variable artificial line in series with the circuit whose equivalent is to be determined. By **cutting out** sections of the artificial line, the total loss in the circuit is made the same as that in the calibrating circuit, so that the d-c meter gives the same deflection in both cases. The dials are arranged to read the loss in the unknown circuit directly.

For transmission measurements at higher frequencies up to 3500 kc, both comparison type and

direct-reading sets are extensively used in the telephone plant. The principles involved are not essentially different from those already discussed for measurements at voice-frequencies, although the measuring sets themselves are necessarily somewhat more elaborate in design. The comparison type sets generally employ thermocouple detectors to drive a direct-reading meter. The receiving circuits of the direct-reading sets are essentially superheterodyne detectors, the outputs

Sending Set

Receiving Set
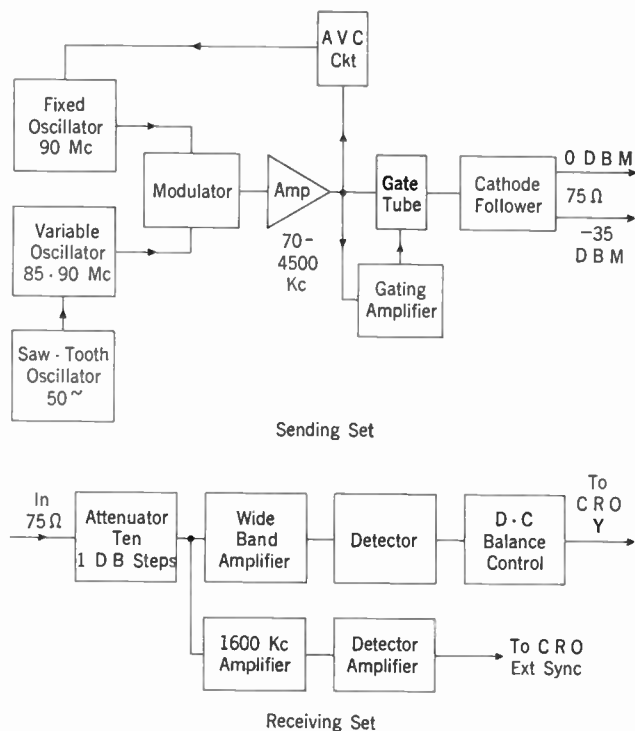
FIG. 33-17 VISUAL TRANSMISSION MEASURING SET

of which are fed to d-c milli- or microammeters reading directly in dbm. Appropriate types of variable oscillators must of course be employed with each measuring set.

In practical transmission maintenance work, it is often necessary to check transmission losses over a considerable range of frequencies to insure that a line is properly equalized to provide uniform transmission over the entire frequency range of the signals to be transmitted. This is particularly true in the case of television signals, where attenuation distortion must be carefully guarded against. Such "attenuation-frequency runs" can be made with most of the types of measuring sets

[356]

discussed above, but the process would obviously be extremely laborious if separate measurements had to be made over a frequency range up to about 3500 kc and the frequency-attenuation curve plotted. To obviate this difficulty, a visual type of transmission measuring set may be used, which plots the attenuation over the entire frequency range under study almost instantaneously on the screen of a cathode-ray oscilloscope. The arrangement of such a transmission measuring set is shown in block diagram in Figure 33-17.

The sending circuit here is designed to apply to the line under test 1 milliwatt of power over a range of frequencies sweeping linearly with time between 70 kc and 4500 kc, at a repetition rate of 50 times per second. As the drawing indicates, the sweep is developed from a heterodyne oscillator in which the 90 megacycle output of a fixed oscillator is combined in a modulator circuit with the output of a variable oscillator, which sweeps linearly over the range 85 to 90 mc. The linear sweep of the variable oscillator is controlled by a 50-cycle saw-tooth oscillator whose output controls the effective capacitance of a reactance tube in the resonant circuit of the variable oscillator. The modulator output includes the sum and difference products of the two oscillators. The sum is in a very high frequency range and is heavily attenuated, while the difference product between zero and 5 mc is transmitted to the amplifier, which has a flat gain over the range 70 to 4500 kc.

The amplifier output is supplied to a "gate tube" which is normally merely a wideband amplifier. A portion of the amplifier output is also supplied to an automatic volume control circuit which holds the net amplifier output constant over the frequency range. Another portion is supplied to the "gating amplifier". The purpose of this device is to prevent frequencies at, or in the immediate neighborhood of, the L-carrier pilot frequencies from reaching the line where they would naturally tend to upset the line-up of the whole system. The gating amplifier circuit contains sharply resonant circuits tuned to the pilot frequencies, by means of which it is able to recognize when the sweeping frequency of the sending set is approaching a pilot frequency. Upon such recognition, it delivers a "squelching pulse" to the gate tube, which effectively blocks transmission through that tube's circuit of a narrow band of frequencies extending about 50 cycles on each side of the pilot frequency. The output of the gate tube is delivered to the line through an impedance-matching cathode-follower circuit. The output power is normally held at 1 milliwatt, although a 35 db pad may be switched into the circuit to provide a −35 dbm output when a lower level of output power is desirable.

The receiving circuit includes an attenuator, a wideband amplifier and a germanium crystal detector circuit whose output is connected through a d-c balance control circuit to the Y plates of a cathode-ray oscilloscope. The purpose of the d-c balance control circuit is merely to center the pattern on the oscilloscope screen. The receiving circuit also includes a synchronizing circuit for keeping the horizontal time sweep of the oscilloscope in phase with the oscillator sweep of the sending circuit. For this purpose, a frequency of 1600 kc (any other single frequency in the band would do equally well) is picked off in front of the wideband amplifier by an amplifier tuned sharply to 1600 kc. This is rectified in the detector-amplifier to apply a synchronizing pulse to the EXT SYNC connection of the oscilloscope each time the received sweeping input passes through 1600 kc.

As each frequency over the 70 to 4500 kc range is received, a d-c voltage proportional to its amplitude is produced by the receiving circuit and applied to the Y axis of the oscilloscope. When the X axis of the oscilloscope is swept at the same rate as that of the sending sweep oscillator, a point is plotted for each frequency and a continuous line is therefore drawn (except for the very narrow bands suppressed at the pilot frequencies). If attenuation over the whole frequency range is uniform, the oscilloscope pattern will be a straight horizontal line; irregularities in the straight line pattern will indicate corresponding deviations in the attenuation at frequencies which may be easily read from the oscilloscope scale. Such deviations may be rapidly corrected by adjusting the appropriate equalizer circuits while observing the pattern on the oscilloscope screen.

33.6 Measurements at Superhigh Frequencies

Figure 33-18 is a simplified schematic of a frequency-power meter that may be used to measure frequency and power levels in the range between 3600 and 4600 megacycles. As may be noted, the meter circuit includes three principal elements— a waveguide system shown at the upper left, a temperature controlling oscillator circuit below this, and an indicating circuit at the right.
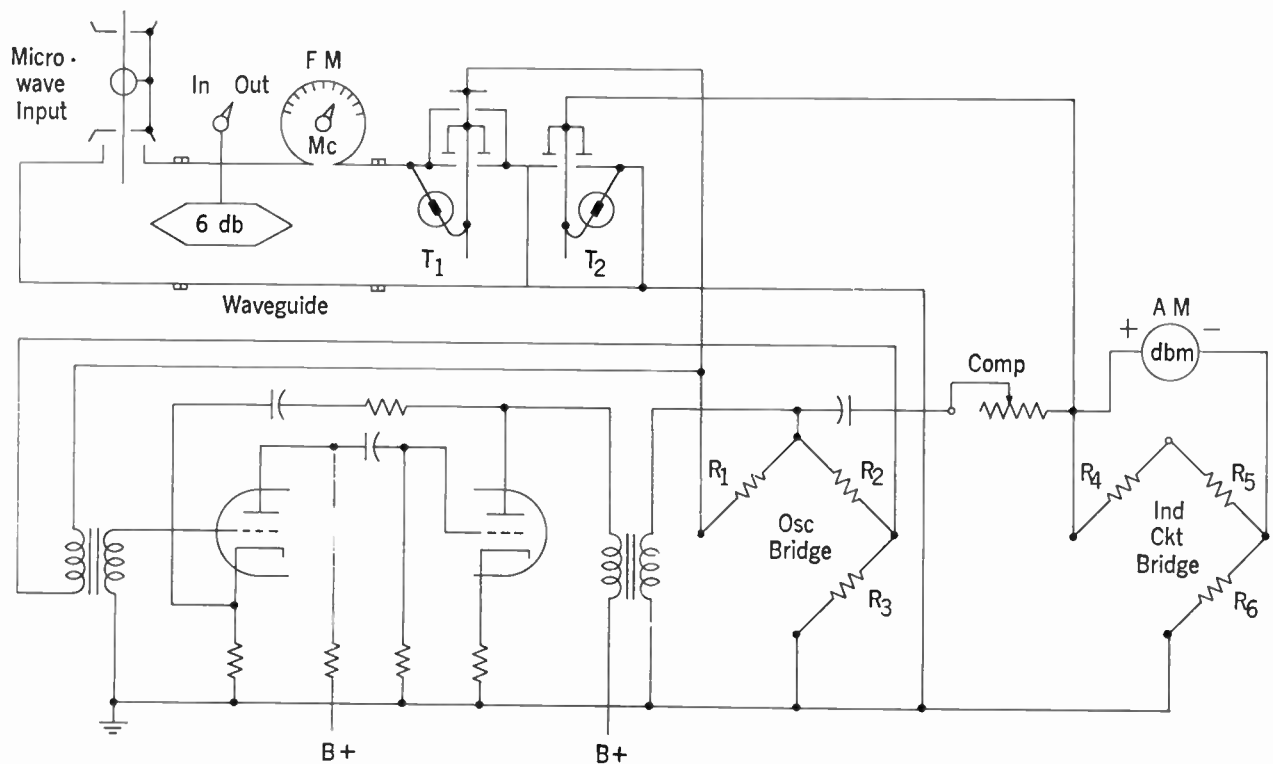
FIG. 33-18 MICROWAVE FREQUENCY-POWER METER

Microwave power is measured by the amount of heat the waves generate in a resistive load consisting of the thermistor bead T_1 indicated at the right of the waveguide system. Frequency is measured by means of the adjustable cavity resonator FM which is coupled to the guide and calibrated to read directly in megacycles. As the movable plunger of the cavity is adjusted to resonance, the reading of the power meter AM will show a sharp dip as a result of most of the energy being absorbed in the cavity and not reaching the thermistor bead. The setting of the dial which mechanically controls the movable plunger then indicates the frequency directly. The input microwave power can be reduced up to 6 db, if desired, by means of an attenuator which consists of a vane of insulating material with a thin layer of carbon on one side that may be moved toward the center of the waveguide by a calibrated control knob.

The thermistor T_1, which is the power measuring element, serves also as a matched resistance load. Its resistance must therefore be held at a constant value in order to properly terminate the waveguide. Since the resistance of a thermistor changes sharply with temperature variations, this means that the temperature of T_1 must be precisely held. The temperature controlling oscillator performs this function, and the microwave power absorbed by the thermistor is determined indirectly by measuring the change in oscillator output power that is required to hold the thermistor's temperature constant.

The oscillator, which operates at 85 kc, is of the bridge-stabilized type. It is comprised of a 2-stage amplifier stabilized by negative feedback, with its output connected to its input through a bridge as shown. Thermistor T_1 forms one arm of the bridge and the other three arms are fixed resistors, R_2 being equal to R_3. The resistance of R_1 is approximately equal to that of T_1 when the thermistor bead is at its normal control temperature. When the oscillator is first turned on, the thermistor is below its normal temperature and the bridge is unbalanced. This causes a large positive feedback current to flow, which in turn increases the oscillator output voltage and hence the current through T_1. The temperature of the thermistor accordingly increases and its resistance decreases until the bridge approaches perfect

balance and a stable condition is reached. The resistance and temperature of the thermistor bead then remains constant as long as the oscillator continues in operation. Any tendency of the temperature of the thermistor to change as a result of the absorption of microwave power is immediately counteracted by an offsetting change in the oscillator output.

The temperature controlling oscillator will also hold the temperature of T_1 constant despite changes in ambient temperature to which the thermistor bead is naturally exposed. In doing so, however, the oscillator output is affected by ambient temperature changes, as well as by microwave power input. The oscillator output cannot, therefore, be used as a direct measure of the microwave power input. To overcome this difficulty, a second "compensating thermistor" T_2 is employed. This thermistor is electrically and thermally similar to T_1 and is so mounted as to be subjected to the same external temperature conditions.

The compensating thermistor is connected across the oscillator output transformer through a capacitor and the COMP potentiometer and is therefore also heated by the oscillator output. Although it is not in the oscillator positive feedback circuit like T_1, proper adjustment of the COMP potentiometer will cause both thermistors to be held at the same fixed temperature as long as no microwave input power is supplied.

When microwave power is sent into the waveguide system, only thermistor T_1 is affected since T_2 is shielded from the incoming microwaves. To prevent the temperature of T_1 from changing, the oscillator output decreases in proportion to the incoming microwave power. This, however, decreases the current in T_2 as well as in T_1 and the temperature of T_2 accordingly decreases proportionately. The resultant proportional increase in the resistance of T_2 unbalances the indicating circuit bridge whose fixed resistance arm R_6 is equal to the normal control resistance of T_2 (R_4 and R_5 being equal fixed resistors). The unbalance causes a flow of current in the microammeter which is directly proportional to the change in resistance of T_2 and hence to the microwave power reaching T_1. The meter is calibrated to indicate the microwave power directly in dbm.

INDEX